TRANSFER EFFECTS OF FUNCTION-BASED WORKING MEMORY TRAINING

Thesis
presented to the Faculty of Arts and Social Sciences
of the University of Zurich
for the degree of Doctor of Philosophy

by Carla De Simoni

Accepted in the spring semester 2017 on the recommendation of
Prof. Dr. Klaus Oberauer (main supervisor)
and
Prof. Dr. Florian Schmiedek (co-supervisor)

Zurich, 2019

Abstract

The enhancement of cognitive abilities through training has attracted broad interest during the last years. It is based on the idea of overlapping processes, which means that the improvement of a cognitive ability elicits enhancement in abilities that share common underlying processes. Working memory (WM) is often targeted for training interventions as it is related to a broad range of other higher-order cognitive abilities such as reasoning and executive functions. Some previous studies, indeed found transfer effects to reasoning. However, the growing number of training studies who found null effects, questioned the robustness of transfer effects. Therefore, the two studies in this thesis investigated the specific mechanisms that drive training and transfer, and an effect that might counteract them.

The first study of this thesis was conducted to examine the impact of WM training on WM capacity and efficiency by evaluating near transfer to untrained WM tasks and far transfer to closely related abilities (i.e., reasoning, processing speed, task switching, and inhibitory control). Moreover, we were also interested in process-specific effects of updating and binding training on three WM mechanisms (i.e., focus switching, removal of WM contents, and interference resolution). We randomly assigned 196 young adults to one of three training interventions (updating, binding, or an active control who trained visual search tasks). Before and after five weeks of adaptive training, performance was assessed with a large test battery measuring all the cognitive processes and abilities. All three training groups revealed evidence for enhanced performance on the tasks they practiced. However, even though WM correlated at least moderately with the transfer measures, we found consistent evidence for the lack of training-induced improvements across all ranges of transfer. The absence of transfer effects indicates that neither WM capacity nor efficiency of specific WM mechanisms was improved through training. Instead, the results suggest that participants acquired task- and material-specific strategies during the training intervention. An alternative explanation for the

absence of transfer, are ego-depletion effects that might have counteracted training and transfer effects.

Therefore, in the second study, we investigated whether the engagement in multiple cognitively demanding tasks lead to ego-depletion and, thus, diminished performance. Ego-depletion refers to a state of impaired self-control, which emerges after acts of self-control. According to the strength model, self-control depends on a limited resource that is exerted by use. However, recent failed replication attempts and meta-analyses raised skepticism about the robustness of ego-depletion effects. Hence, we reanalyzed four previously reported data sets from effortful cognitive test batteries, to test two predictions derived from the strength-model: 1) ego-depletion yields a performance decline over time, 2) ego-depletion effects are shown for all cognitive abilities requiring self-control. However, with few exceptions Bayesian hypothesis tests provided evidence in favor of null effects, resulting in the rejection of the hypotheses. Thus, the current reanalysis show that the strength model does not contribute to our understanding of performance in cognitive tasks.

Zusammenfassung

Die Verbesserung von kognitiven Fähigkeiten durch Training stiess während den letzten Jahren auf grosses Interesse. Es basiert auf der Idee von überlappenden Prozessen, was bedeutet, dass die Steigerung einer kognitiven Fähigkeit Verbesserungen in nahe verwandten Funktionen, die grundlegende gemeinsame Prozesse teilen, hervorruft. Arbeitsgedächtnistraining wird oft für solche Trainingsinterventionen verwendet, da es mit vielen anderen kognitiven Funktionen wie schlussfolgerndes Denken und Exekutive Funktionen stark korreliert. Tatsächlich haben einige frühere Studien einen Transfer auf schlussfolgerndes Denken gefunden. Allerdings führt die stetig wachsende Anzahl von Trainingsstudien, welche Null-Effekte finden, zu Skepsis hinsichtlich der Häufigkeit von Transfer. Die folgenden zwei Studien untersuchten spezifische Mechanismen, die Training und Transfer hervorrufen und ein Effekt, welcher ihnen möglicherweise entgegenwirkt.

In der ersten Studie dieser Arbeit untersuchten wir die Auswirkungen von Arbeitsgedächtnistraining auf die Kapazität und Effizienz von Arbeitsgedächtnis, indem wir nahen Transfer auf ungeübte Arbeitsgedächtnisaufgaben und fernen Transfer auf eng verwandte Fähigkeiten (d.h. Intelligenz, Verarbeitungsgeschwindigkeit und Exekutive Funktionen) massen. Darüber hinaus interessierten uns prozessspezifische Effekte von Updating- und Binding-Training auf drei Arbeitsgedächtnismechanismen (die Fähigkeit den Fokus der Aufmerksamkeit zu verschieben, das Entfernen von nicht länger relevanten Arbeitsgedächtnisinhalten, und die Überwindung von Interferenz).

Wir haben 196 junge Erwachsene zufällig eine von drei Trainingsinterventionen zugewiesen (Updating, Bindung oder aktive Kontrolle, die visuelle Suchaufgaben trainiert hat). Vor und nach fünf wöchigem Training wurden die Leistung in den relevanten kognitiven Fähigkeiten mittels einer grossen Testbatterie erhoben. Alle drei Trainingsgruppen zeigten Evidenz für eine verbesserte Leistung in den von ihnen geübten Aufgaben. Obwohl

Arbeitsgedächtnis zumindest moderate Korrelationen mit den Transfermassen zeigte, fanden wir deutliche Evidenz für das Fehlen einer trainingsbedingten Leistungssteigerung in den Transferaufgaben. Das Fehlen von Transfereffekten deutet darauf hin, dass weder die Kapazität noch die Effizienz von spezifischen Arbeitsgedächtnismechanismen durch Training verbessert werden konnten. Stattdessen lassen die gefunden Ergebnisse vermuten, dass die Teilnehmer während des Trainings aufgaben- und materialspezifische Strategien erworben haben. Eine alternative Erklärung für das Fehlen von Transfer sind Ego-Depletion-Effekte, die den Trainings- und Transfer-Effekten entgegengewirkt haben könnten.

In der zweiten Studie haben wir deshalb untersucht, ob das Bearbeiten einer Serie von mehreren kognitiv anspruchsvollen Aufgaben zu Ego-Depletion und damit verbundenen Leistungseinbussen führt. Ego-Depletion bezieht sich auf einen Zustand verminderter Selbstkontrolle, welcher nach Handlungen, die ein hohes Mass an Selbstkontrolle benötigen, auftaucht. Entsprechend dem Modell der Selbstkontrollstärke hängt die Selbstkontrolle von einer begrenzten Ressource ab, die bei Gebrauch reduziert wird. Jüngste fehlgeschlagene Replikationsversuche und Metaanalysen erzeugten Zweifel hinsichtlich der Beständigkeit von Ego-Depletion-Effekten. Daher haben wir Datensätze von vier bereits durchgeführten kognitiven Testbatterien neu analysiert. Basierend auf dem Modell der Selbstkontrollstärke haben wir die zwei Hypothesen formuliert und getestet: 1) Ego-Depletion führt über die Zeit zu einem Leistungsabfall, 2) Ego-Depletion-Effekte werden für alle kognitiven Fähigkeiten, die Selbstkontrolle erfordern, gezeigt. Mit wenigen Ausnahmen lieferten Bayes'sche Hypothesentests Evidenz für Nulleffekte, was zur Ablehnung der beiden Hypothesen führte. Die aktuellen Reanalysen zeigen, dass das Modell der Selbstkontrollstärke nicht in der Lage ist, die Leistung in kognitiven Aufgaben zu erklären.

Content

1 Introduction

The potential of computerised cognitive training interventions has become a hotly debated topic in the last decade. Brain trainings (e.g., Lumosity, Cogmed, CogniFit, Brain Age, Posit Science InSight and BrainFit) which claim to improve cognitive abilities such as working memory (WM), executive functions and reasoning enjoy great popularity and serve a rapidly expanding market. The market research firm SharpBrains (www.sharpbrains.com), which is specialized for health and wellness applications of brain science predicted yearly sales for cognitive training software of $3.38 billion by 2020 (SharpBrains, 2013, 2015). Moreover, they estimated that about half of the purchases of training software are made by schools, employers or health providers, with the other half made by consumers (SharpBrains, 2015). Despite the commercial success of such training interventions, empirical evidence about its effectiveness is still inconclusive. This thesis investigates the specific mechanisms that underlie training gains and an effect that might counteract training and transfer gains.

1.1 Theory behind training

The rationale behind cognitive training interventions is that the training of one ability generalize to untrained cognitive abilities when underlying cognitive processes are shared (cf. Jaeggi et al., 2010; Klingberg, 2010). This assumption is based on early work of transfer conducted by Thorndike and colleagues, who argued that transfer occur when the trained and transfer function share common elements (cf. Thorndike & Woodworth, 1901). Moreover, imaging studies showed that overlapping processes engage in the same brain regions and hence, transfer from one function to another could be expected (cf. Buschkuehl, Jaeggi, & Jonides, 2012; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008). For example, Dahlin and colleagues (2008) found that training of a letter-memory task engaging on striatal regions generate transfer to an n-back task, which is associated to the same region, but not to a Stroop task, which tap another brain region.

Working memory (WM) is the ability that it is most often targeted by such training interventions (cf. Morrison & Chein, 2011; Shipstead, Hicks, & Engle, 2012; Shipstead, Redick, & Engle, 2010). WM is a system that allow for temporary access to information that are needed for a current task (von Bastian & Oberauer, 2014). It refers to the ability to integrate, update, and maintain information (cf. Oberauer, 2009; Oberauer & Hein, 2012; Oberauer, Süss, Wilhelm, & Sander, 2007; Wilhelm, Hildebrandt, & Oberauer, 2013). Individual differences in WM have been linked to a series of higher-order cognitive abilities, such as reasoning (Friedman, Miyake, Schmeichel, & Tang, 2006; Kyllonen & Christal, 1990; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Süß, Oberauer, Wittman, Wilhelm, & Schulze, 2002), executive functions (Friedman et al., 2009; Miyake et al., 2000a; Miyake & Friedman, 2012), reading comprehension (de Jonge & de Jong, 1996), mental arithmetic (DeStefano & LeFevre, 2004), and academic achievement (St Clair-Thompson & Gathercole, 2006). Moreover, impairments in WM are related to numerous neurological disorders such as the attention deficit hyperactivity disorder (ADHD; e.g., Martinussen, Hayden, Hogg-Johnson, & Tannock, 2005). According to the theory about overlapping processes, training of WM was assumed to not only enhance closely related functions (e.g., reasoning), but also to reduce WM impairments induced by neurological disorders (cf. Klingberg, 2010; von Bastian & Oberauer, 2014). Early training interventions, in which participants trained WM task across several weeks revealed transfer effects to reasoning (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg, Forssberg, & Westerberg, 2002). However, after initial euphoria about this seminal finding there is now growing skepticism about the impact of WM training interventions on related abilities evoked by an increasing number of studies, which could not replicate transfer effect after working memory training (Brehmer, Westerberg, & Bäckman, 2012; Chein & Morrison, 2010; L. Dunning, Holmes, & E. Gathercole, 2013; Redick et al., 2013; von Bastian & Eschen, 2016). Recent meta-analyses on this topic revealed that training interventions mainly lead to an improvement in the practiced task. Evidence for transfer effects (e.g., to reasoning) is more mixed, and if found rather small (Au et al., 2015; Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Schwaighofer, Fischer, & Bühner, 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; Weicker, Villringer, & Thöne-Otto, 2015). However, these findings have to be interpreted cautiously due to persisting methodological issues in the field of working memory training. Recent reviews on cognitive training research (cf. Green, Strobach, & Schubert, 2014; Morrison & Chein, 2011; Noack, Lövdén, Schmiedek, & Lindenberger, 2009; Shipstead et al., 2012, 2010; von Bastian & Oberauer, 2014) discuss primarily the lack of a theory driven selection of training and transfer tasks, the assessment of cognitive abilities with single tasks, the lack of an adequate control group, and small sample sizes.

1.2 How to design a training intervention

The above mentioned reviews on cognitive training research gave best practices on how to design a training intervention. First, they highly recommend to base the selection of training and transfer tasks on established theories (cf. von Bastian & Oberauer, 2014). Because only then one can attribute potential transfer effects to a specific ability or cognitive process. In general, cognitive improvements after training can be induced by two processes, namely increased WM capacity or by a more efficient allocation of the capacity available (for a detailed discussion see von Bastian & Oberauer, 2014). Increased capacity should target performance in all tasks that depend on the capacity-limit. Hence broad transfer effects could be expected in case of expanded capacity. In contrast, efficiency could either be increased by higher automatization of core processes (e.g., by speeding-up specific processes) or by the acquisition of strategies. Whereas, the former would lead to relatively large transfer (i.e., to abilities that drive on such core processes), the latter would rather generate narrow transfer to similar tasks (cf. von Bastian & Oberauer, 2014). As mentioned, to identify the process that is responsible for training and transfer effects, a training intervention has to be based on a WM model that defines limits for both aspects. Hence, we based our thesis on the three-embedded-components model of WM (cf. Oberauer, 2009; Oberauer & Hein, 2012).

The three-embedded-components model (Oberauer, 2009; Oberauer & Hein, 2012) rests upon Cowan's WM model (Cowan, 1995). It distinguishes three functional levels of information selection namely the activated part of long-term memory (aLTM), the region of direct access (RDA), and the focus of attention (FoA). The aLTM comprises all the information that is needed for a current task. Perceptual input activates representations in long-term memory, which then activate other related representations through spread of activation. For example, when solving an arithmetic problem (e.g., "8 - 5 = 3"), digits and math operators will be activated in long-term memory. A subset of these activated representations is held in the capacity-limited RDA, where they are temporarily combined into new bindings. In regard to the mentioned arithmetic equation, this means that the elements of the equation are connected to their respective roles (e.g., 8 is the first operand, "-" is the operator). The RDA is the equivalent of the focus of attention in Cowan's WM model (1995). In contrast, the FoA in Oberauer's model, selects one item held in the RDA that is relevant for current processing (e.g., the digit "8" of the equation).

Second, each ability should be measured with multiple indicators. Previous cognitive training studies often measured ability with only one task (cf. von Bastian & Oberauer, 2014). However, performance on a single task is not a perfect indicator of an ability, but also reflects ability-independent variance, caused by systematic and random influences (cf. Shipstead, Redick, & Engle, 2012, Noack, Lövden, Schmiedek, 2014, Green, Strobach, Schubert, 2013). On the other side cognitive abilities could not explain the overall variance of a single indicator tasks. Thus, if the aim is to make statements about ability-related improvements, constructs should be measured with multiple heterogeneous tasks targeting the same ability (cf. Lövden, 2014). However, measuring each ability with multiple indicators has the negative side effect that the duration of test batteries for pre-and posttest is considerably prolonged. Green and colleagues (2014) argued that the engagement in test batteries that last several hours could lead to ego depletion (cf. Baumeister, 2014; Baumeister & Heatherton, 1996) alongside with severe performance decline and hence might cover potential transfer effects. The idea behind the concept of ego-depletion is originated in the strength-model of self-control developed by Baumeister and colleagues (Baumeister, 2014; Baumeister & Heatherton, 1996; Baumeister, Vohs, & Tice, 2007). According to this model, self-control draws on a limited resource that diminishes by use. Self-control exertion generates a temporary state of "ego-depletion" during which further acts of self-control are more prone to failure. Performance declines are assumed to aggravate the less of the resources is available (Baumeister, 2014; Vohs, Baumeister, & Schmeichel, 2013). Moreover, the resources is shared among various processes, thus performance impairments are assumed to occur in a wide range of behaviors including WM and executive functions (Schmeichel, 2007; Webb & Sheeran, 2003). For example, in a study by Schmeichel (2007) participants first watched a short video-clip, whereby half of the participants were instructed to ignore distractions on the bottom of the screen (depletion condition) and the other half to just watch the clip (control condition). Afterwards participants had to complete a complex span task, in which they had to alternating memorized a target and complete a distractor task. After an unpredictable number of memoranda and distractor tasks, participants had to recall the memoranda. Results showed that participants of the depletion condition performed significantly worse in the complex span task than participants of the control condition. Schmeichel (2007) concluded that ignoring the on-screen distractions exerted resources and hence lead to impaired performance in the WM task. Hence, performance in cognitive test batteries administered before and after a training intervention might be highly influenced by ego-depletion effects (cf. Green et al., 2014).

Third, there are still a considerable number of studies that compare the effects of their training intervention to a passive control group (cf. Dougherty, Hamovitz, & Tidwell, 2015; Melby-Lervåg et al., 2016; von Bastian & Oberauer, 2014). Passive control groups participate in pre- and posttest but do not underwent a training intervention, hence this group could only control for test-retest effects (cf. Shipstead et al. 2012). However, to additionally account for intervention effects such as regularly practicing computer-based tasks, an active control group is needed (von Bastian & Oberauer, 2014). Indeed meta-analyses showed that transfer effects are overestimated when effects of a training intervention are compared to a passive control group (Dougherty et al., 2015; Melby-Lervåg et al., 2016, but see Au et al., 2015). Consequently, Simons et al. (2016) denote the inclusion of an active control group as the gold standard in training research. The ideal control training intervention should be as similar as possible to the experimental training in terms of contextual conditions and design. It should only differ in the ability that is trained (von Bastian & Oberauer, 2014). In contrast to a pharmacological study in which participants do not know whether they are taking the effective drug or the ineffective placebo, participants of a training study are not completely blind to the condition, as they directly experience what training they are undergoing (Boot, Simons, Stothart, & Stutts, 2013). The mere expectancy that a training intervention could improve cognitive functions, could already be sufficient to cause performance increase at posttest (Morrison & Chein, 2011). In order to convincingly assign training success to an intervention, it is therefore essential that participants of both groups have comparable expectancies regarding the training outcome (Boot et al., 2013). However, there is still no consensus about the ideal active control group (von Bastian & Oberauer, 2014). In the empirical training study included in this thesis, we decided to employ visual search tasks as control intervention. These tasks have already been successfully used in previous training studies (cf. Harrison et al. 2013, von Bastian, Langer, Jäncke, & Oberauer, 2013) and show only weak correlations to WM (e.g. Kane, Poole, Tuholsky, & Engle, 2011).

Finally, another methodological issue that not only concerns training research, but psychological research in general, are the small sample sizes that lead to low statistical power (i.e., the probability of finding genuine effects is small). For example, in the meta-analysis by Melby-Lervåg et al. (2016) the average sample size of the 87 included studies is slightly above 20 participants per group. Small sample sizes enhance the probability of false-negative, false-positive, and inflated effect-size estimates (Button et al. 2013). For example, Melby-Lervåg et al. (2016) found the largest effect size (d =2.18) for transfer to reasoning in a training study by Klingberg et al. (2002) which only included 7 participants per group. Furthermore, Melby-Lervåg and colleagues showed that training induced improvements in reasoning are only observed in studies with small sample sizes (less than 20 participants per group) and untreated control group (Melby-Lervåg et al., 2016). Therefore, it could be expected that a large number of effect sizes are overestimated (see also von Bastian, Guye, & De Simoni, 2017 for a detailed discussion). To circumvent the described power issues, we recruited at least 59 participants per group for out training intervention. Moreover, we decided to use Bayesian inference as an alternative approach to evaluate the strength of evidence. In this approach, strength of evidence is expressed by the Bayes factor (*BF*) which describes the probability of the data in one hypothesis (e.g., cognitive training yield transfer to related abilities) in comparison to the probability of the data in another hypothesis (e.g., training does not yield transfer).

1.3 Empirical Studies

The goal of the first study reported in this thesis (*No Evidence for Effects of Updating and Binding Training on Working Memory Capacity and Efficiency*), was to evaluate the effectiveness of WM training to elicit near transfer to untrained, structurally different WM tasks and far transfer to closely related cognitive functions (i.e., reasoning, processing speed, task switching and inhibitory control), and to examine the specific WM mechanisms (i.e., focus switching, removal of WM contents, and interference resolution) that potentially drive these effects. For this purpose, we randomly assigned 196 young adults to two experimental groups, one of which underwent a memory updating training and the other an associative binding training, and to an active control group, who practiced visual search tasks. Cognitive processes and abilities of interest were assessed with a large test battery that was administered before and after five weeks of intense training. In this test battery each of the cognitive abilities was measured with four tasks covering verbal-numerical and visuo-spatial material. Training and transfer data were analyzed with Bayesian linear mixed-effect (LME) models. Furthermore, we conducted structural equation models to investigate relationships among the assessed cognitive abilities. Results showed that all three training group improved in the tasks they trained. However, even though WM revealed at least moderate correlations to the transfer measures, there was compelling evidence for the absence of any training induced transfer. These findings suggest that participants could neither improve WM capacity nor efficiency of specific WM mechanisms through training, but instead might have acquired task-specific strategies. However, another potential reason for the absence of transfer could be the occurrence of ego-depletion effects during pre- and posttest. Ego-depletion effects alongside with performance impairments are assumed to emerge when a series of cognitively demanding tasks have to be carried out (e.g., Vohs, Baumeister, & Schmeichel 2013). The test battery administered at pre- and posttest consisted of 28 cognitive tasks and lasted about five hours. As Green et al. (2014) argued, transfer effects (which tend to be rather small when found, e.g. Soveri et al. 2017), might have been obscured by such ego-depletion effects. To investigate potential ego-depletion effects across multiple cognitive tasks, we reanalyzed data of four previous studies including large cognitive test batteries.

Therefore, in the second study reported in this thesis (*No Consistent Evidence for Ego-Depletion Effects across Multiple Cognitive Tasks and Domains*) we investigated whether the engagement in multiple cognitively demanding tasks lead to a performance decline, as proposed by the strength-model of self-control (Baumeister, 2014; Baumeister & Heatherton, 1996; Baumeister et al., 2007). Hence, we reexamined four studies where participants were engaged in test batteries of cognitively demanding tasks for several hours, to test the following two hypotheses: 1) ego-depletion yields a performance decline over time, 2) ego-depletion effects are task-independent and affect all cognitive abilities requiring self-control. The test batteries were administered in two versions, with half of the participants completing the test battery in exact reversed order. To test the two hypotheses, we first carried out one-tailed two sample Bayesian t-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) to compare performance on each task in forward and backward order. Based on the strength-model, we assumed that participants in the forward order outperformed those in the backward order in the first half of the test battery and vice versa. Moreover, to accumulate evidence across the four data sets, we conducted small-scale meta-analyses (cf. Cumming, 2014) with a Bayesian approach. With few exceptions, all of the Bayesian hypothesis tests provided evidence in favor of null effects, thus both hypotheses were rejected. These findings render it unlikely that ego-depletion effects diminished training and transfer gains.

2 No Evidence for Effects of Updating and Binding Training on Working Memory Capacity and Efficiency

Carla De Simoni[1], and Claudia C. von Bastian[2]

[1] University of Zurich, Department of Psychology

[2] Bournemouth University, Department of Psychology

Submission status

Submitted for publication.

Authors' contributions

CDS: Review of the literature, development of the research question and goals, design of the study, development of the test battery and training tasks, programming the test battery with Tatool, data collection, data analysis and interpretation, writing of the manuscript

CvB: Supervision and discussion of CDS's contributions, revising the manuscript

Acknowledgement

This work was supported by a grant from the Swiss National Science Foundation (No. 100014_146074) to Klaus Oberauer. Moreover, both authors were supported by the URPP "Dynamics of Healthy Aging" of the University of Zurich. We thank Clara Overkott, Samuel Winiger, Aline Gaillard, Doris Pfister, Monalisa Stiefel, and Aylin Yantaz for their assistance with collecting the data.

Abstract

The effectiveness of working memory (WM) training is still under debate with previous research yielding inconsistent findings. In this study, we aimed to disentangle effects on WM capacity and efficiency by evaluating not only near transfer to untrained, structurally different WM tasks and far transfer to closely related abilities (i.e., reasoning, processing speed, task switching, and inhibitory control), but also process-specific effects of updating and item-context binding training on three WM mechanisms (i.e., focus switching, removal of WM contents, and interference resolution). We randomly assigned 196 young adults to one of two experimental groups (updating or binding) or to an active control group practicing visual search tasks. Before and after five weeks of adaptive training, performance was assessed measuring each of the cognitive processes and abilities of interest with four tasks covering verbal-numerical and visual-spatial materials. All three training groups revealed evidence for an improvement on the tasks they practiced. However, despite the relatively large sample size and at least moderate correlations between WM and the transfer measures, we found consistent evidence for the absence of training-induced improvements across all ranges of transfer. The absence of any transfer indicates that neither WM capacity nor efficiency of specific WM mechanisms was improved during training. Instead, the results suggest that participants developed task- and material-specific strategies through training.

Keywords

working memory, capacity and efficiency, updating, item-to-context binding, cognitive training

2.1 Introduction

During the last decade, computer based working memory (WM) training has received considerable attention. After the initial enthusiasm sparked by early empirical evidence for generalization of training gains to untrained abilities (e.g., fluid intelligence, cf. Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg, Forssberg, & Westerberg, 2002; von Bastian & Oberauer, 2013), there is now a growing body of studies suggesting that the beneficial effects of working memory training may have been overestimated (e.g., Harrison et al., 2013; Sprenger et al., 2013; von Bastian & Eschen, 2016). Potential reasons for these inconsistencies lie in the wide variety of sometimes flawed research methods used to assess WM training and transfer gains (cf. von Bastian & Oberauer, 2014). Recent meta-analyses (Au et al., 2015; Dougherty, Hamovitz, & Tidwell, 2015; Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Schwaighofer, Fischer, & Bühner, 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; Weicker, Villringer, & Thöne-Otto, 2015) yielded moderate near transfer effects (i.e., transfer to untrained WM tasks) and small, sometimes significant far transfer effects (i.e., transfer to abilities related to WM, such as reasoning and executive functions). These meta-analytic findings, however, have to be interpreted cautiously, as they are based on studies with low statistical power due to small sample sizes, with the average treatment group consisting of about 20 participants (e.g., Melby-Lervåg et al., 2016). Low statistical power increases the probability of not only false-negative and false-positive findings (e.g., Button et al., 2013), but can also inflate effect sizes. For example, in their simulation study, Halsey, Curran-Everett, Vowler, and Drummond (2015) showed that attempts to detect a true medium effect (Cohen's d = 0.50) with low statistical power (n = 30, theoretical power = 48 %) yielded 97 % of inflated effects sizes, with significant effect sizes ranging from $d = 0.44$ to $d = 1.23$. As the true size of transfer effects is unknown, we can only speculate about the number of inflated effect sizes in training research; however, meta-analytic effect size estimates are likely to be overestimated. Another approach to evaluate the evidence of training and transfer effects is the use of Bayesian inference, where the strength of evidence is expressed by the Bayes factor (*BF*). The *BF* is the likelihood of the data under one hypothesis (usually the alternative hypothesis, H_1) relative to the likelihood of the data under the other hypothesis (usually the null hypothesis, H_0, cf. Jeffreys, 1961). In contrast to null hypothesis statistical testing (NHST), Bayes statistics allows for quantifying evidence not only for the alternative hypothesis (i.e., the presence of training and transfer effects) but also for the null hypothesis (i.e., absence of training and transfer effects). Consequently, Bayes statistics are increasingly popular in cognitive training research (e.g., Dougherty et al., 2015; Guye, De Simoni, & von Bastian, 2017; Guye & von Bastian, 2017; Sprenger et al., 2013; von Bastian & Oberauer, 2013). For example, Dougherty and colleagues (2015) recently reevaluated the 20 n-back training studies included in the meta-analysis by Au et al. (2015) with a Bayesian approach. Out of the 24 comparisons, 11 (i.e., 46 %) contributed only ambiguous evidence ($BF < 3$), indicating that the data from these studies were not sensitive enough to support either hypothesis. Given that the average group size in the included studies was only $n = 20$, the ambiguity of the results was probably due to low power. Hence, evidence regarding training and transfer effects of WM is still inconclusive, with large-scale training studies contributing stronger evidence still being needed.

Besides methodological issues, much training research also suffered from a lack of theory-driven investigations. Generally, training-induced cognitive improvements can be caused by two processes: increased WM capacity or enhanced WM efficiency (for a more detailed discussion, see von Bastian & Oberauer, 2014). Increased capacity is reflected in structural changes and should therefore lead to broad transfer manifested in improved performance across tasks drawing on this capacity-limit. In contrast, enhanced efficiency refers to a more efficient exploitation of the capacity available, for example through using task-specific strategies or a higher level of automatization of specific WM mechanisms. Whereas strategy-use can be expected to transfer only to very similar tasks (e.g., Morrison & Chein, 2011), a higher level of automatization should result in transfer to tasks that draw on the same WM mechanisms (cf. von Bastian & Oberauer, 2014). However, the specific mechanisms of working memory (WM) that are potentially responsible for training and transfer effects have received only little attention yet, with few exceptions (Gibson, Gondoli, Johnson, Steeger, & Morrissey, 2012; Hussey et al., 2016; Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013; Waris, Soveri, & Laine, 2015).

To disentangle the effects of enhanced capacity and enhanced efficiency, we based our current work on the three-embedded-components model of working memory (Oberauer, 2009; Oberauer & Hein, 2012). This model is an extension of the model proposed by Cowan (1995) and assumes three functional levels of information selection, namely the activated part of long-term memory (aLTM), the region of direct access (RDA), and the focus of attention (FoA). The aLTM reflects all representations needed for a current task, activated through perceptual input or spread of activation. In the RDA, a subset of the activated representations is temporarily bound into new structures. In contrast to the aLTM, the capacity of the RDA is limited due to interference between simultaneously maintained bindings (Oberauer, 2005). Lastly, the FoA selects the one item of the RDA that is processed next. According to the three-embedded components model, increased WM capacity would result in an increased number of bindings that can be maintained at a time. As the binding hypothesis assumes that the capacity to build and maintain temporary bindings is the common limiting factor that explains the strong correlation between working memory and reasoning (Oberauer, Süss, Wilhelm, & Sander, 2007), increased WM capacity in terms of a larger number of simultaneously maintained bindings should, hence, be reflected in broad transfer to reasoning and other cognitive abilities (e.g., executive functions) potentially drawing on the same mechanism of simultaneously maintaining bindings.

Better WM performance could also reflect a more efficient use of basic WM mechanisms such as enhanced focus switching, removal of outdated information, and interference resolution. Focus switching refers to the ability of the FoA to flexibly shift between single items held in the RDA. Improved focus switching is reflected by a reduction of the time needed to move the FoA from one item to another, increasing the time to refresh memoranda (Barrouillet, Bernardin, & Camos, 2004), thus potentially yielding enhanced recall performance. There is some evidence showing that the cost in reaction times (RT) induced by switching the FoA can be decreased by training (e.g., Dorbath, Hasselhorn, & Titz, 2011; Oberauer, 2006; Verhaeghen, Cerella, & Basak, 2004).

Removal of outdated information is defined as the „unlearning or unbinding of an item from its context" (Ecker, Lewandowsky, & Oberauer, 2014, p. 3). More efficient removal is reflected by a reduction in the time to remove no longer relevant information from WM. Oberauer & Lewandowsky (2016) argued that removal is a basic operation of working memory. For the proper functioning of working memory, especially the building of new bindings in the RDA, it is essential that outdated information is removed, because otherwise it would strongly interfere with the information that is relevant for a current task.

Finally, interference resolution describes the ability to overcome interference among bindings held in the RDA. This WM mechanism becomes important whenever a conflict in information processing occurs. For example, in a recognition task, conflict occurs when participants are presented with recent material that however was not part of the current memory set; hence the material is highly familiar yet presently irrelevant. In such cases, there is a strong tendency to respond yes (cf. Hussey et al., 2016). Resolve this interference requires the recollection of the item and its context (i.e., the whole binding, Oberauer, 2005; see also Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011). Enhanced interference resolution is reflected by improved recollection performance. Au et al. (2015) assume that interference resolution is one of the most central processes responsible for the transfer effects found after n-back training (cf. Oelhafen et al., 2013).

Present Study

In the present study, our goal was to evaluate the effectiveness of WM training to elicit near and far transfer effects, and to examine the specific mechanisms that potentially drive these effects. For this purpose, we compared two training interventions, memory updating training and associative binding training, to an active control group. The application of updating training is widespread in cognitive training research (cf. Soveri et al., 2017; von Bastian & Oberauer, 2014) and updating tasks have often been successfully used in training interventions (e.g., *n-back tasks:* Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Lilienthal et al., 2013; Redick et al., 2013, *keep-track tasks*: Dahlin, Neely, Larsson, Backman, & Nyberg, 2008; Dahlin, Nyberg, Bäckman, & Neely, 2008, *running-memory tasks*: Waris et al., 2015, and *memory updating tasks*: Linares, Borella, Lechuga, Carretti, & Pelegrina, 2017; Florian Schmiedek, Lövdén, & Lindenberger, 2010). Updating training can be assumed to put strong demands on focus switching (cf. Oberauer, 2006) and on removal of outdated information (cf. Ecker et al., 2014; Ecker, Lewandowsky, Oberauer, & Chee, 2010), and, to a lesser degree, on interference resolution.

In addition to updating training, we also evaluated associative binding training. Based on the binding hypothesis (Oberauer et al., 2007), a training regimen targeting the ability to simultaneously maintain multiple bindings can be assumed to maximize chances to observe broad transfer to reasoning. Associative binding tasks put strong demands on interference resolution (e.g., Oberauer, 2005), which has been hypothesized to potentially mediate transfer effects (cf. Au et al., 2015; Jaeggi et al., 2010). Still, process-based training interventions only rarely target associative binding directly, with the few existing studies having focused on older adults (Bellander et al., 2017; Zimmermann, von Bastian, Röcke, Martin, & Eschen, 2016). Moreover, associative binding training may also affect the efficiency of removal and, to a lesser degree, focus switching.

Besides the specific WM mechanisms, we were also interested in far transfer to abilities that have been shown to strongly correlate with WM, such as reasoning (Friedman, Miyake, Schmeichel, & Tang, 2006; Kyllonen & Christal, 1990; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Süß, Oberauer, Wittman, Wilhelm, & Schulze, 2002), shifting and inhibition (Friedman et al., 2009; Miyake et al., 2000; Miyake & Friedman, 2012), and processing speed (Schmiedek, Oberauer, Wilhelm, Süss, & Wittmann, 2007).

Recent reviews and meta-analyses gave best practice recommendations for training interventions (cf. Green, Strobach, & Schubert, 2014; Melby-Lervåg et al., 2016; Noack, Lövdén, & Schmiedek, 2014; Shipstead et al., 2012; Simons et al., 2016; von Bastian & Oberauer, 2014), such as the inclusion of an active control group, the usage of multiple indicators to measure cognitive abilities, and an adequate sample size. We considered these best practices in our study design. First, we included an active control group to differentiate between effects that emerge due to the training intervention and those caused by participating in a study (cf. von Bastian & Oberauer, 2014). Participants of the control group trained visual search tasks, which put only weak demands on WM (e.g., Kane, Poole, Tuholski, & Engle, 2006; Sobel, Gerrie, Poole, & Kane, 2007) and were successfully used in previous training studies (e.g., Harrison et al., 2013; Redick et al., 2013; von Bastian, Langer, Jäncke, & Oberauer, 2013). Second, we assessed each cognitive function with four indicators to prevent task-specific features being responsible for the observation of potential transfer effects (cf.

Noack et al., 2014; Shipstead et al., 2012). Third, our training groups comprised between 59 and 72 participants, thus our group sample sizes were about three times as large as the size of average treatment groups in cognitive training research (i.e., n = 20, cf. Soveri et al., 2017).

2.2 Method

Participants completed 20 training sessions of extensive cognitive training over the course of 5 weeks. We randomly assigned participants to either one of two experimental groups (i.e., updating or binding training) or an active control group that underwent visual search training. The study followed a double-blind design, hence neither the participants nor the experimenter conducting pre- and posttest knew to which group participants were assigned to. We assessed training and transfer effects with a large test battery before and after training. The same test battery was used at both assessments to facilitate between-groups baseline comparisons, which is needed for investigating the comparability across groups and occasions.

2.2.1 Participants

We recruited young adults for a "cognitive training study" through the participant pool of the Department of Psychology of the University of Zurich, postings at the university campus, and a short study presentation in lectures. The experimental protocol was approved by the institutional review board at the University of Zurich. All participants were German native speakers or highly proficient in German and signed consent to participate. Moreover, participants reported normal or corrected-to-normal vision, no color blindness or current psychiatric or neurological disorders, and no psychotropic drug use. All participants gave written consent to taking part in the study. As previous studies were likely severely underpowered (cf. Bogg & Lasecki, 2015) and, hence, probably yielded inflated effect size estimates (Halsey et al., 2015), we refrained to use those estimates for power analyses. Instead, we aimed to recruit at least three times as many participants than previous studies (i.e., n = 60 per group). We managed to recruit 240 participants of which 219 completed the study. Reasons for the 21 dropouts were lack of time (7), loss of interest (1), technical issues (4), or personal issues (1). An additional 8 participants withdrew consent without comment. We had to exclude another 23 participants due to a programming error (10) and lack of compliance (i.e., performance below chance level in more than five (i.e., 25 %) training sessions, (13)). Thus, the final sample included in the statistical analyses consisted of 196 participants. Table 1 lists the descriptive statistics for demographic variables. There was no evidence for group differences in the demographics (BFH0 ≥ 1.17), except for an age difference between the binding and the updating group (BFH1 = 12.56). Participants were reimbursed after posttest completion (CHF 120, or 10 course credits and CHF 20). Moreover, participants received a bonus of up to CHF 50 depending on the level of difficulty they achieved during training (cf. von Bastian & Oberauer, 2013).

Table 1

Participant Demographics

	Group		
Demographics	Updating	Binding	Active Control
Sample size (*n*)	59	65	72
Gender (f/m)	40/19	45/20	49/23
Age (*M* ± SD)	22.61 ± 2.97	24.62 ± 4.14	23.79 ± 4.17

Note. BFs indicated support for the null hypothesis that there were no group differences as determined by Bayesian Pearson chi-square test (gender) and Bayesian two-tailed *t*-tests (age).

2.2.2 Design and Materials

Training. Participants completed training at their own computer or laptop at home using Tatool (von Bastian, Locher, & Ruflin, 2012), a Java-based open-source training and testing tool (www.tatool.ch). After each session, data were automatically uploaded to a web server running Tatool Online. An experimenter, who was not involved in the collection of outcome measures, monitored participants' training performance and served as contact person during training. As in previous work (von Bastian & Oberauer, 2013; von Bastian et al., 2013; von Bastian & Eschen, 2016), we tried to maximize experimental control through automated online analyses to detect irregularities (e.g., performance below chance level). To increase individual commitment, participants (1) signed a participation agreement at the beginning of the study, (2) were made aware of their progress being constantly monitored, (3) received regular emails (i.e., after 2 and 4 weeks of training) on their training progress, and (4) were reminded to practice when falling behind their training schedule (i.e., less than four sessions completed per week).

Each group practiced four tasks with varying material (numerical, verbal, visual, and spatial, see Figure 1 and Table 2) for approximately 10 min each per training session. The order of task administration was randomized for each training session. Task duration was restricted to a maximum of 11.25 min so that a training session did not exceed 45 min. The updating group completed up to 12 trials, the binding group up to 24 trials, and the active control group up to 100 trials per task and session.

Updating training. In the four memory updating tasks (adapted from Lewandowsky, Oberauer, Yang, & Ecker, 2010), participants had to remember an initial set of simultaneously presented stimuli. During the updating phase, participants had to transform individual stimuli (e.g., mentally move previously memorized circles in a grid or applying a simple arithmetic operation to a digit), enter the result of the transformation, and remember the result. In half of the trials, a cue indicated which of the stimuli had to be updated next. After nine updating steps, participants had to recall the most recent result of each stimulus.

Binding training. In the four associative binding tasks (adapted from Oberauer, 2005; Wilhelm, Hildebrandt, & Oberauer, 2013), participants had to memorize sequentially presented associations of two elements (e.g., symbol and digits or fractals and their location). In the subsequent recognition phase, each association was randomly probed with one of the elements given as cue. Overall, 50 % of the probes were positive (i.e., exact matches), 25 % were intrusions (i.e., probes that were presented in the current trial, but associated with a different element), and 25 % were distractors (i.e., probes not presented in the current trial).

Active control training. In the four visual search tasks (von Bastian et al., 2013), participants had to decide whether a target (e.g., a single-headed arrow) was present in a search display consisting of horizontally and vertically presented distractors (e.g., double-headed arrows), and if so, to indicate its orientation (up, down, left, or right) by pressing the corresponding arrow key. In target-absent trials, participants had to press "A" instead.

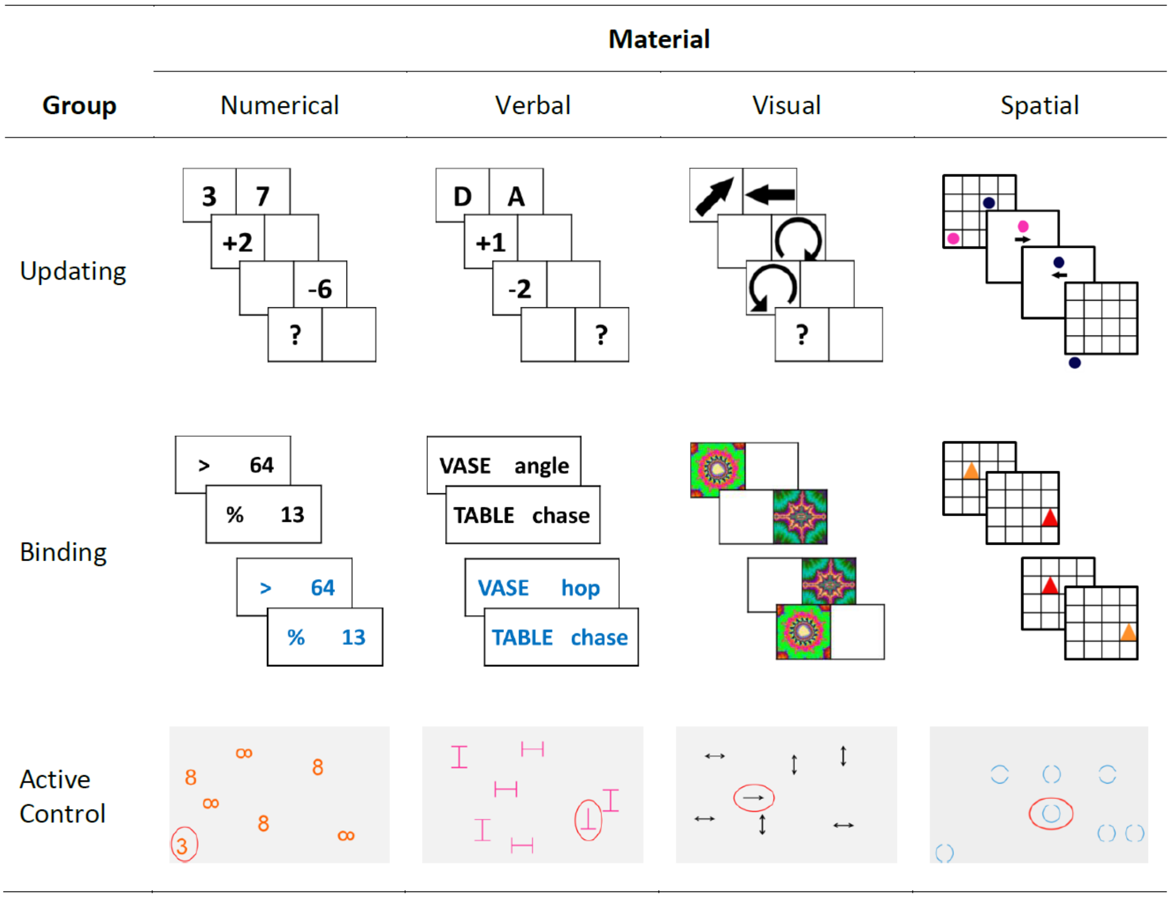

Figure 1. Schematic depiction of the four training tasks in each intervention (see text for details).

Adaptive training algorithm. Each training task's difficulty was adjusted to individual performance. Updating and binding training started with a set size (i.e., number of memoranda) of two and a maximal response time limit of 3500 ms, and active control training with a search display containing six stimuli. The first training session served to evaluate participants' individual cognitive performance limit (cf. Guye & von Bastian, 2017). In each task, participants' mean accuracy was measured after every 10 % of trials (1 trial in the updating tasks, 2 trials in the binding tasks, and 10 trials in the visual search tasks). If participants scored at least 85 % correct, difficulty was increased, otherwise it remained on the current level. For the remaining training session (i.e., sessions 2-20) participants' performance was checked after every 40% of trials (5 trials in the updating tasks, 10 trials in the binding tasks, and 40 trials in the visual search tasks). In the updating and binding tasks, difficulty was adjusted by decreasing the response time limit by 500 ms for four subsequent level-ups (e.g., when reaching levels 2 through 5), and by increasing set size by one additional memorandum every fifth level-up (e.g., when reaching level 6). The response time limit was reset to 3500 ms whenever the set size was increased and again reduced by 500 ms for the subsequent four level-ups. In the active control group, level-ups corresponded to an increase in set size (i.e., number of stimuli shown in the search display) by one additional stimulus. We refrained from adjusting response time in the active control group, to minimize demands on processing speed, which is strongly correlated with WM and reasoning (cf. Schmiedek et al., 2007). Participants were informed when they reached a higher difficulty level (i.e., “Congratulations, you achieved the next level”), and they started each session on the highest level they achieved in the previous session (cf. von Bastian et al., 2013).

Questionnaires. Participants were asked to monitor their learning progress during training. Therefore, they were asked two questions at the beginning of each training session (i.e., "how would you rate your performance in the last training session?", "how well will you perform in today's training session?") and one at the end of each session (i.e., "how would you rate your performance in today's training session?"). These data will be reported elsewhere. After each training session, participants additionally answered a set of questions about the enjoyment experienced and the effort spent during the training session, and the perceived fit between difficulty and ability (see von Bastian & Eschen, 2016). Furthermore, participants completed the German version of the Questionnaire on Current Motivation (QCM, Rheinberg, Vollmeyer, & Burns, 2001) after training sessions 1 and 10, and the Intrinsic Motivation Inventory (IMI, Deci & Ryan, 2015) after training session 20. These data will be reported in Guye, De Simoni, and von Bastian (2017). After training session 20, participants completed a questionnaire in which they were asked whether they used strategies to complete the tasks, and, if so, to rate how helpful they were on a scale from 1 (*not at all)* to 5 (*very*).

Pre- and posttest. Before and after the training intervention, practice, near, and far transfer effects were assessed with a test battery comprising 28 computer-based tasks. Each cognitive ability was assessed with four tasks with varying material (see Table 2 for an overview). Up to four participants were tested simultaneously in a single lab session that lasted about 5 h, including three short breaks. To control for fatigue and practice effects, half of the participants in each group completed the test battery in reversed order (i.e., they started with the forward's order last task of the last block and finished the test battery with the first task of the first block, cf. von Bastian & Oberauer, 2013). To familiarize participants with the tasks, several practice trials were presented before test blocks of pseudorandomized trials.

Before pretest, participants had to complete the following personality questionnaires: NEO-FFI (Borkenau & Ostendorf, 2008); Need for Cognition (Bless, Wänke, Bohner, Fellhauer, & Schwarz, 1994), Theories of Intelligence Scale (Dweck, 2000), Grit Scale (Duckworth & Quinn, 2009), Self-Efficacy Scale (Schwarzer & Jerusalem, 1995), Self-Efficacy to Regulate Exercise (Bandura, 2006). Findings from these measures are reported in Guye et al. (2017).

Training. Training tasks were administered as criterion tasks at pre- and posttest to compare practice effects between training conditions. The proportion of correctly recalled items served as dependent variable for the updating tasks. The discrimination parameter *d*' from signal detection theory served as outcome measure for the associative binding tasks. It was computed from subtracting the *z*-transformed false alarms to intrusion probes from the *z*-transformed hit rates (cf. Oberauer, 2005). Participants' ability to search through a visual display was measured by the individual residuals from a simple linear regression model predicting the RTs of target absent trials from RTs of target present trials (for a similar approach in dual tasks see Oberauer, Lange, & Engle, 2004).

WM mechanisms. To investigate the WM mechanisms (i.e., focus switching, removal speed, and interference resolution) targeted in the two experimental training interventions, we assessed three outcome measures. *Focus switching* was measured by object-repetition benefits costs in the updating tasks (cf. Oberauer & Hein, 2012) by subtracting RTs to switch trials from RTs to repetition trials. *Removal speed* was measured with four modified memory updating tasks (cf. Ecker et al., 2014). As in the updating tasks, participants had to memorize an initial set of stimuli. In the subsequent updating phase, however, individual stimuli were substituted with new ones and participants had to press the space bar (or click a button in the locations version of the task, cf. Table 2) as soon as they had memorized the new stimulus. A cue that was presented either 200 ms (i.e., short cue-target interval, CTI) or 1500 ms (i.e., long CTI) indicated which stimulus was updated next. After 1 to 18 updating steps, participants had to recall the most recent stimuli. Individual residuals from a simple regression model predicting the RTs of short CTI from RTs of long CTI were used as dependent variables (cf. Ecker et al., 2010). Finally, *interference resolution* was gathered by the proportion of correct responses to intrusion probes in the binding tasks. Moreover, we used the proportion of correct responses to matching and distractor probes in binding tasks as additional outcome measures for explorative analyses.

Near Transfer. To assess near transfer, we investigated whether participants of the updating group showed transfer to binding tasks and vice versa. The discrimination parameter *d*’ computed from the binding tasks and the mean recall accuracy of the updating tasks were used as dependent variables.

Far Transfer. Transfer to related cognitive abilities (i.e., reasoning, shifting, processing speed, and inhibition) were measured with four tasks. The four *reasoning* tasks required participants to either detect a rule behind a pattern or to integrate information for drawing a conclusion. The proportion of correctly answered items served as outcome measure. In the four *shifting* tasks, participants had to categorize bivalent stimuli according to one of two classification rules as indicated by a cue (cf. von Bastian, Souza, & Gade, 2016). Each task consisted of five blocks: two single-rule blocks (only one rule had to be applied, e.g., animacy classification followed by size classification), a mixed-rules block (two rules switched, e.g., switching between animacy and size classifications), and another two single-rule blocks in reversed order (e.g., size classification followed by animacy classification). Switching costs were calculated by subtracting RTs to switch trials from RTs to repetition trials in mixed-rules block. *Processing speed* was gathered by average RTs in single-rule blocks of the shifting tasks. In the four *inhibition* tasks, participants were required to inhibit prepotent responses. These tasks comprised three conditions: a congruent condition (correct and prepotent response correspond), an incongruent condition (correct and prepotent response do not correspond) and a neutral condition (no prepotent response present). Interference scores, which were computed by subtracting the RTs to incongruent trials from RTs to neutral trials, served as output measures.

Table 2

Description of the Training and Transfer Tasks

Task (Material)	Description	Conditions (%)	No. Trials	Set Size	Display Duration (ms)
	Updating				
Digits (n)	Memorize a list of digits (1-8) simultaneously presented in a row of boxes. Subsequently, update stimulus by a simple arithmetic operation (ranging from -7 to + 7) presented in one of the boxes, enter the result, and memorize the new digit in that box. After 9 updating steps, recall the most recent digit in each box (adapted from Lewandowsky et al., 2010).	50 switch 50 repetition	18	3-5	stimulus: 500 cue: 500 operation: u.r.
Letters (ve)	Memorize a list of letters (A-H) simultaneously presented in a row of boxes. Subsequently, update stimulus by mentally shifting it 1-3 positions forward or backward in the alphabet, enter the result, and memorize the new letter in that box. After 9 updating steps, recall the most recent letter in each box (adapted from Lewandowsky et al., 2010).	50 switch 50 repetition	18	2-4	stimulus: 500 cue: 500 operation: u.r.
Arrows (vi)	Memorize a list of arrows (8 arrows each 45 degrees apart from the next one when arranged in a circle) simultaneously presented in a row of boxes. Subsequently, update stimulus by mentally rotating them 45 degrees clockwise or counter clockwise, enter the result and memorize the new direction of the arrow. After 9 updating steps, recall the most recent results in each box (adapted from Lewandowsky et al., 2010).	50 switch 50 repetition	18	2-4	stimulus: 500 cue: 500 operation: u.r.
Locations (s)	Memorize the position of a set of colored circles (8 different colors) simultaneously presented in a 4x4 grid. Subsequently update the position of the circles by mentally shifting them one cell up, down, left or right, indicate the new position in a blank grid, and remember the new location of the circle. After 9 updating steps, recall the most recent positions of each circle (adapted from Lewandowsky et al., 2010).	50 switch 50 repetition	18	3-5	stimulus: 500 cue: 500 operation: u.r.
	Binding				
Symbol-Digit (n)	Memorize sequentially presented pairings of mathematical symbols and two digit numbers. In the subsequent recognition phase, decide whether the probe pairs match those during memorization (adapted from Wilhelm et al., 2013).	50 positive 25 intrusion 25 new	24	4-6	stimulus: 900 probe: u.r.
Noun-Verb (ve)	Memorize sequentially presented noun-verb pairs. In the subsequent recognition phase, decide whether the probe pairs match those during memorization (adapted from Wilhelm et al., 2013).	50 positive 25 intrusion 25 new	24	4-6	stimulus: 1800 probe: u.r.

Fractal-Location (vi)	Memorize the location of fractals sequentially presented in a row of boxes. In the subsequent recognition phase, decide whether the locations of the probe fractals match those during memorization (adapted from Oberauer, 2005).	50 positive 25 intrusion 25 new	24	4-6	stimulus: 900 probe: u.r.
Color-Location (s)	Memorize the location of colored triangles sequentially presented in a 4 x 4 grid. In the subsequent recognition phase, decide whether the color of the probe triangles match those during memorization (adapted from Oberauer, 2005).	50 positive 25 intrusion 25 new	14	4-6	stimulus: 1800 probe: u.r.
	Visual Search				
Numbers (n)	Decide whether a search display contains a 3 among horizontally and vertically presented 8's, and if so indicate its direction (up, down, left or right; adapted from von Bastian et al., 2013).	80 present 20 absent	30	7,9,11[b]	u.r.
Letters (ve)	Decide whether a search display contains a capital T among horizontally and vertically presented I's, and if so indicate its direction (up, down, left, or right; adapted from Harrison et al., 2013).	80 present 20 absent	30	7,9,11[b]	u.r.
Arrows (vi)	Decide whether a search display contains a single headed arrow among horizontally and vertically presented double-headed arrows, and if so indicate its direction (up, down, left or right; adapted from von Bastian et al., 2013).	80 present 20 absent	30	7,9,11[b]	u.r.
Circles (s)	Decide whether a search display contains a circle with one gap among circles with two gaps, and if so indicate its direction (up, down, left or right; adapted from von Bastian et al., 2013).	80 present 20 absent	30	7,9,11[b]	u.r.
	Removal				
Digits (n)	Memorize a list of digits (1-8) simultaneously presented in a row of boxes. In the subsequent updating phase, a cue indicates which box has to be updated next. Remove the outdated information from memory and press the space bar after having memorized the new digit presented in that box. After 1 to 18 updating steps, recall the most recent digit in each box (adopted from Ecker et al., 2010).	50 long CTI 50 short CTI	18	3	stimulus: 500 long cue: 1800 short cue: 200 new item: u.r.
Letters (ve)	Memorize a list of letters (A-H) simultaneously presented in a row of boxes. In the subsequent updating phase, a cue indicates which box has to be updated next. Remove the outdated information from memory and press the space bar after having memorized the new letter presented in that box. After 1 to 18 updating steps, recall the most recent letter in each box (adopted from Ecker et al., 2010).	50 long CTI 50 short CTI	18	3	stimulus: 500 long cue: 1800 short cue: 200 new item: u.r.
Arrows (vi)	Memorize a list of arrows (8 arrows each 45 degrees apart from the next one when arranged in a circle) simultaneously presented in a row of boxes. In the	50 long CTI 50 short CTI	18	3	stimulus: 500 long cue: 1800

	subsequent updating phase, a cue indicates which box has to be updated next. Remove the outdated information from memory and press the space bar after having memorized the direction of the new arrow presented in that box. After 1 to 18 updating steps, recall the most recent direction of the arrow in each box (adopted from Ecker et al., 2010).				short cue: 200 new item: u.r.
Locations (s)	Memorize the position of a set of colored circles (8 different colors) simultaneously presented in a 4 x 4 grid. In the subsequent updating phase, a cue indicates which circle position has to be updated next. Remove the outdated information from memory and click a button below the grid after having memorized the new position of the circle. After 1 to 18 updating steps, recall the most recent position of each circle (adopted from Ecker et al., 2010).	50 long CTI 50 short CTI	18	3	stimulus: 500 long cue: 1800 short cue: 200 new item: u.r.
	Reasoning				
Diagramming Relationships (ve)	Choose the diagram that best describes the relationship between three nouns. Complete as many items as possible in 8 min (cf. Ekstrom, French, Harman, & Dermen, 1976).		30	5[a]	u.r.
Letter Sets (ve)	Find the logical pattern underlying sets of four letters to identify the deviating letter set. Complete as many items as possible in 14 min (cf. Ekstrom et al., 1976).		30	5[a]	u.r.
Locations Test (vi)	In four rows of dashes, one dash is replaced by an “x” according to a particular rule. Infer the position of the “x” in the fifth row. Complete as many items as possible in 12 min (cf. Ekstrom et al. 1976).		28	5[a]	u.r.
RAPM (s)	Select the piece that best fills the gap in a matrix pattern. Complete as many items as possible in 15 min (adapted from Raven, 1990).		28	8[a]	u.r.
	Shifting				
Parity – Magnitude (n)	Decide whether a digit between 1 and 9 (excluding 5) is either odd or even, or smaller or larger than 5 (von Bastian et al., 2016).	50 switch 50 repetition	385	8[b]	u.r.
Animacy – Size (ve)	Classify sketches of animals and objects according to their animacy (living or non-living) or to their size (smaller or larger than a soccer ball, von Bastian et al., 2016).	50 switch 50 repetition	385	64[b]	u.r.
Fill – Frame (s)	Decide whether a geometric figure is dotted or striped, or framed or not (adapted from von Bastian et al., 2016).	50 switch 50 repetition	385	64[b]	u.r.
Color – Shape (vi)	Categorize figures according to their color (blue or green) or their shape (round or angular, von Bastian et al., 2016).	50 switch 50 repetition	385	64[b]	u.r.
	Inhibition				

Numerical Stroop (n)	Decide which of two digits has the larger value. In congruent trials the value and the size of the digit match (i.e., the digit with the higher value is displayed larger), in incongruent trials there is a discrepancy between value and size (e.g., the digit 2 is displayed larger than the digit 3). In neutral trials both digits have the same size (adapted from Tzelgov, Meyer, & Henik, 1992).	33 congruent 33 incongruent 33 neutral	288	24[b]	u.r.
Color Stroop (ve)	Indicate the ink (blue or green) of color words (blue or green). In congruent trials the ink matches the color word (e.g., word BLUE written in blue), but not in incongruent trials (e.g., word BLUE written in green). In neutral trials, colored Xs are presented (adapted from Stroop, 1935).	33 congruent 33 incongruent 33 neutral	288	12[b]	u.r.
Global-Local (s)	Indicate the local shape (circle or square) of a global figure (circle or square) or the shape of a line-drawing. In congruent trials the local shape is identical, in incongruent trials it is different to the global shape. In neutral trials, line drawings of circles or squares are presented (adapted from Navon, 1977).	33 congruent 33 incongruent 33 neutral	288	6[b]	u.r.
Simon (vi)	Press the left arrow key when a green circle is displayed and the right arrow key when a red circle is displayed on the screen. In congruent trials, the location of the circle and the location of the response key match (e.g., a green circle is presented on the left), but not in incongruent trials (e.g., a green circle is presented on the right). In neutral trials, the circle is displayed in the center of the screen (adapted from Simon, Sly, & Vilapakkam, 1981).	33 congruent 33 incongruent 33 neutral	288	6[b]	u.r.

Note. All tasks included two practice trials (12 in the shifting and inhibition tasks) to familiarize participants with the tasks. Set size reflects the number of memoranda. U.r. = until response; ve = verbal, vi = visual, n = numerical, s = spatial.

[a] Number of response options.

[b] Number of different stimuli.

2.3 Analyses

First, we were interested in whether we could confirm the proposed theoretical transfer model in the pre- and posttest data by using structural-equation modeling. Second, we analyzed the training data to check whether participants improved on the practiced tasks and how motivated they were during training with Bayesian linear mixed-effects (LME) models. Third, we investigated gains from pre- to posttest on the level of constructs using again Bayesian LME models. Finally, we examined whether participants used any strategies during training.

2.3.1 Structural-Equation Modeling

To investigate the relationship between the measured cognitive abilities at pre- and posttest, we conducted a latent-variable confirmatory factor analyses (CFA) using the "lavaan" package (Rosseel, 2012). We examined model fit evaluating the chi-square statistic (χ^2), the comparative fit index (CFI), the root mean square error of approximation (RMSEA) and its 90% CI, and the standardized root mean-squared residual (SRMR). Established cut-off values indicating good fit are values above .95 for CFI, values less than 0.06 for RMSEA, and values below .08 for SRMR (Hu & Bentler, 1999).

2.3.2 Bayesian Linear Mixed-Effects Modeling

We used Bayes statistics for the analyses of the training and transfer data. BFs range on a continuous scale from 0 to ∞, with a BF of 1 reflecting perfect ambiguity (i.e., the data support both hypotheses equally). BFs below 1 represent evidence for the hypothesis in the denominator (typically H0), and a BF above 1 evidence in favor of the hypothesis in the numerator (typically H1). For example, a BF of 10 in favor of the H1 means that the data are ten times more likely under H1 than H0.

All analyses were conducted in R (R Core Team, 2015) with the “BayesFactor” package (Morey & Rouder, 2015). We used the default prior settings (i.e., Cauchy distribution with a scaling factor r = 0.707) to compute Bayes Factors and the verbal classifications suggested by Wetzels and Wagenmakers (2012, Table 3) for their interpretation. Bayes Factors below 1 (i.e., BFs in favor of the H0) are expressed as 1/BF to facilitate interpretation.

Table 3

Verbal Labels for Interpreting Bayes Factors

Bayes Factor	Label
1 to 3	Anecdotal
3 to 10	Substantial
10 to 30	Strong
30 to 100	Very strong
> 100	Decisive

Note. Adapted from Wetzels and Wagenmakers (2012)

We conducted Bayesian LME models using the lmBF function of the Bayes Factor package. These models have the advantage that they simultaneously account for multiple sources of variance in the data. Two types of effects are distinguished in LME models, fixed effects (e.g., variance from experimental conditions or predictors) and random effects (e.g., variance from individual differences). In the present study, we included participant and tasks as crossed-random effects (Baayen, Davidson, & Bates, 2008), which accounts for the fact that both participants and tasks included in our study are random samples drawn from larger populations.

First, we evaluated whether baseline cognitive performance was comparable across groups by running Bayesian LME models on the level of constructs including performance as dependent variable, group as predictor, and participant and task as crossed-random factors.

Second, we examined training performance for each training task separately by conducting a Bayesian LME using the maximum load achieved in each training session as dependent variable, session as fixed effect, and participants as random factor. Training session was coded as linear contrast to investigate monotonic trends across sessions. Third, we evaluated motivation during training by running a Bayesian LME with training motivation measures (enjoyment, effort and perceived fit between task difficulty and ability; cf. von Bastian & Eschen, 2016) as dependent variable, group and session as fixed factors, and participant as random effect. Fourth, we investigated training and transfer gains from pre- to posttest. Thus, we computed standardized gain scores (i.e., posttest performance minus pretest performance divided by pretest standard deviation) for each participant and each task (von Bastian & Oberauer, 2013). We compared performance between each experimental group and the active control group for the abilities measured in pre- and posttest by running separate Bayesian LME models with gain score as dependent variable, group as fixed effect and participants and task as crossed random effects. Finally, we investigated whether participants used certain strategies during training and as how helpful they perceived them. Therefore, we conducted a Bayesian Pearson chi-square test to evaluate whether groups differed in the frequency to use strategies, and a Bayesian ANOVA to investigate whether the groups rated the helpfulness of their strategies differently.

2.3.3 RT Data Treatment

Only RTs of correct responses were analyzed. RTs being 3 median absolute deviations away from the overall median (Leys, Ley, Klein, Bernard, & Licata, 2013) were defined as outliers and excluded from analyses. To reduce positive skew of speed-based outcome measures, we additionally applied a log transformation to RT data. All dependent variables were z-transformed across the three groups. To eliminate variance due to the two different orders of test administration in pre- and posttest, we arbitrarily selected one order as the reference condition and corrected the data of the other order for the mean difference between the two orders for each variable (cf. von Bastian & Oberauer, 2013).

2.3.4 Missing Data

Due to technical difficulties at pretest, data of two tasks is lost for one participant each. In addition, we set data with irregularities (i.e., mean accuracy below guessing probability, mean RTs below 250 ms, and suspicious response patterns such as constant repetition of the same key presses) as missing. This concerned 12 tasks for 23 participants at pretest and 10 tasks for 22 participants at posttest (for an overview see Table A1 in the appendix). Participants with missing data were excluded from analyses on those measures. Six participants had difficulties pursuing their training schedule and, hence, did complete only 19 sessions. For the analyses of training progress and motivation during training, we included only participants with complete training data sets.

2.4 Results

Table 4 lists the descriptive statistics and reliabilities for each outcome measure for each group and time. Correlations between tasks for tasks at pre- and posttest are listed in Table A2 in the Appendix. The data are available on the Open Science Framework (https://osf.io/fy5ku).

Table 4

Task Performance as a Function of Training Group and Time of Assessment

	Group							
	Updating		Binding		Active control		Reliabilities	
Tasks	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Updating								
Digits	.75 ± .16	.80 ± .16	.71 ± .20	.75 ± .19	.72 ± .19	.74 ± .20	.94	.94
Letters	.59 ± .21	.73 ± .17	.49 ± .21	.53 ± .22	.56 ± .19	.57 ± .20	.91	.93
Arrows	.40 ± .16	.49 ± .18	.35 ± .17	.41 ± .19	.37 ± .15	.43 ± .16	.86	.89
Locations	.43 ± .17	.58 ± .19	.39 ± .21	.46 ± .22	.39 ± .19	.46 ± .20	.93	.94
Binding								
Symbol-Digit	1.45 ± .47	1.45 ± .57	1.36 ± .49	1.62 ± .75	1.38 ± .47	1.45 ± .59	.71	.84
Noun-Verb	1.92 ± .65	1.90 ± .75	1.94 ± .66	1.92 ± .71	1.81 ± .62	1.77 ± .72	.81	.83
Fractal-Location	1.33 ± .47	1.45 ± .66	1.38 ± .47	2.20 ± .88	1.37 ± .46	1.43 ± .56	.69	.87
Color-Location	1.62 ± .62	1.90 ± .79	1.55 ± .58	1.97 ± .73	1.52 ± .54	1.64 ± .58	.75	.82
Visual Search								
Numbers	.27 ± .35	.26 ± .28	.30 ± .28	.29 ± .27	.28 ± .32	.23 ± .29	.65	.65
Letters	.33 ± .40	.30 ± .36	.36 ± .41	.35 ± .37	.36 ± .38	.25 ± .34	.28	.40
Arrows	.31 ± .35	.33 ± .34	.33 ± .33	.33 ± .35	.32 ± .35	.30 ± .37	.56	.51
Circles	.29 ± .34	.28 ± .33	.32 ± .30	.35 ± .30	.30 ± .32	.28 ± .32	-.06	-.24
Focus switching								
Digits	-.39 ± .13	-.32 ± .12	-.37 ± .16	-.36 ± .17	-.38 ± .20	-.38 ± .14	.67	.63
Letters	-.37 ± .20	-.30 ± .11	-.38 ± .23	-.40 ± .19	-.37 ± .19	-.40 ± .19	.43	.52
Arrows	-.54 ± .25	-.52 ± .16	-.56 ± .33	-.52 ± .40	-.45 ± .35	-.53 ± .27	.63	.74
Locations	-.51 ± .18	-.55 ± .19	-.55 ± .22	-.55 ± .21	-.53 ± .20	-.54 ± .20	.82	.85
Removal								
Digits	-.03 ± .15	-.01 ± .14	-.03 ± .18	.00 ± .15	.00 ± .14	-.01 ± .15	.83	.61
Letters	-.03 ± .17	-.03 ± .15	.00 ± .19	-.01 ± .16	-.01 ± .18	-.05 ± .17	.78	.72

Arrows	.02 ± .15	.02 ± .10	-.03 ± .41	-.01 ± .21	.05 ± .14	-.01 ± .39	.69	.36
Locations	.00 ± .09	-.01 ± .09	.00 ± .13	.00 ± .08	.00 ± .08	-.01 ± .08	.33	.33
Interference Resolution								
Symbol-digit	.65 ± .14	.67 ± .14	.61 ± .14	.66 ± .21	.61 ± .13	.69 ± .13	.67	.77
Noun-verb	.71 ± .15	.73 ± .14	.72 ± .16	.73 ± .14	.69 ± .17	.72 ± .16	.76	.73
Fractal-location	.59 ± .14	.64 ± .16	.63 ± .14	.71 ± .19	.64 ± .13	.68 ± .14	.66	.79
Color-location	.66 ± .16	.72 ± .18	.64 ± .16	.68 ± .18	.65 ± .16	.72 ± .14	.75	.81
Reasoning								
Relationships	.74 ± .14	.75 ± .13	.70 ± .14	.73 ± .17	.75 ± .13	.77 ± .13	.72	.78
Letter Sets	.72 ± .12	.74 ± .16	.73 ± .14	.74 ± .15	.70 ± .15	.73 ± .14	.79	.83
Locations	.53 ± .16	.59 ± .17	.46 ± .18	.54 ± .18	.50 ± .16	.58 ± .17	.75	.79
RAPM	.59 ± .23	.65 ± .23	.57 ± .25	.62 ± .19	.60 ± .19	.63 ± .18	.73	.66
Shifting								
Parity-Magnitude	-.36 ± .15	-.35 ± .16	-.29 ± .15	-.30 ± .19	-.32 ± .18	-.32 ± .17	.85	.90
Animacy-Size	-.39 ± .16	-.40 ± .13	-.31 ± .15	-.31 ± .15	-.34 ± .17	-.34 ± .16	.84	.82
Fill-Frame	-.27 ± .13	-.29 ± .13	-.21 ± .11	-.25 ± .23	-.26 ± .14	-.25 ± .12	.73	.87
Color-Shape	-.34 ± .18	-.32 ± .16	-.24 ± .16	-.25 ± .14	-.28 ± .15	-.31 ± .16	.85	.83
Speed								
Parity-Magnitude	542 ± 48	516 ± 43	557 ± 83	521 ± 73	543 ± 77	524 ± 75	.99	.99
Animacy-Size	611 ± 55	566 ± 50	622 ± 91	583 ± 86	605 ± 72	571 ± 76	.99	.99
Fill-Frame	515 ± 42	473 ± 37	530 ± 61	487 ± 56	516 ± 62	479 ± 56	.99	.99
Color-Shape	503 ± 48	466 ± 42	514 ± 70	489 ± 68	499 ± 74	483 ± 85	.99	.99
Inhibition								
Numerical Stroop	-.04 ± .04	-.06 ± .04	-.05 ± .07	-.06 ± .04	-.06 ± .05	-.06 ± .05	.74	.59
Color Stroop	-.06 ± .07	-.03 ± .05	-.07 ± .08	-.04 ± .07	-.06 ± .08	-.05 ± .07	.78	.77
Global-Local	-.02 ± .06	-.03 ± .06	-.07 ± .12	-.04 ± .06	-.04 ± .06	-.06 ± .14	.79	.78
Simon	-.03 ± .01	-.03 ± .02	-.03 ± .02	-.03 ± .02	-.03 ± .02	-.02 ± .02	.60	.61

Note. Values denote *M*s ± *SD*s. Updating and reasoning values are proportions of correct responses, binding values are *d*', visual search and removal values are given in residuals, and inhibition, shifting, and focus switching values are difference scores.

2.4.1 Evaluation of the Theoretical Transfer Model

To evaluate the theoretical transfer model, we conducted a latent-variable confirmatory factor analyses. As the manifest inhibition variables did not converge to a latent factor, we had to exclude the inhibition factor from the model. The reduced model including six correlated factors for updating, binding, visual search, reasoning, shifting, and general speed fit the data well at pretest, $\chi^2(233) = 340.23$, $p < .001$, CFI = .96, RMSEA = .05 [.04; .06], SRMR = .06 and at posttest, $\chi^2(233) = 341.36$, $p < .001$, CFI = .96, RMSEA = .05 [.04; .06], SRMR = .06. As depicted in Figure 2, all tasks loaded significantly on their respective factor. Moreover, correlations between the latent factors were significantly positive, except for visual search, which exhibited no significant relation to any other factor. Finally, all latent factors exhibited significant variance.

More specifically, the two working memory factors, updating and binding, were strongly related (coefficient estimate = .88), and both factors showed moderate correlations with reasoning (coefficient estimate = .64 and .52, for updating and binding, respectively) at pretest. The pattern was the same at posttest, but the correlation between updating and binding decreased (coefficient estimate = .77), and the correlations to reasoning increased (coefficient estimate = .83 and .63, for updating and binding, respectively). As expected, the updating, binding, and reasoning factors were moderately related to general speed, with coefficient estimates between -.34 and -.48 at pretest, and between -.30 and -.41 at posttest. Furthermore, updating showed a weak, but significant correlation with shifting at pretest (coefficient estimate = -.19) but not posttest (coefficient estimate = -.13).[1]

[1] Although it would have been desirable to run a model for multiple groups, our attempts to do so led to severe convergence issues, potentially due to the small group sizes. Therefore, only models across participants from all groups are reported.

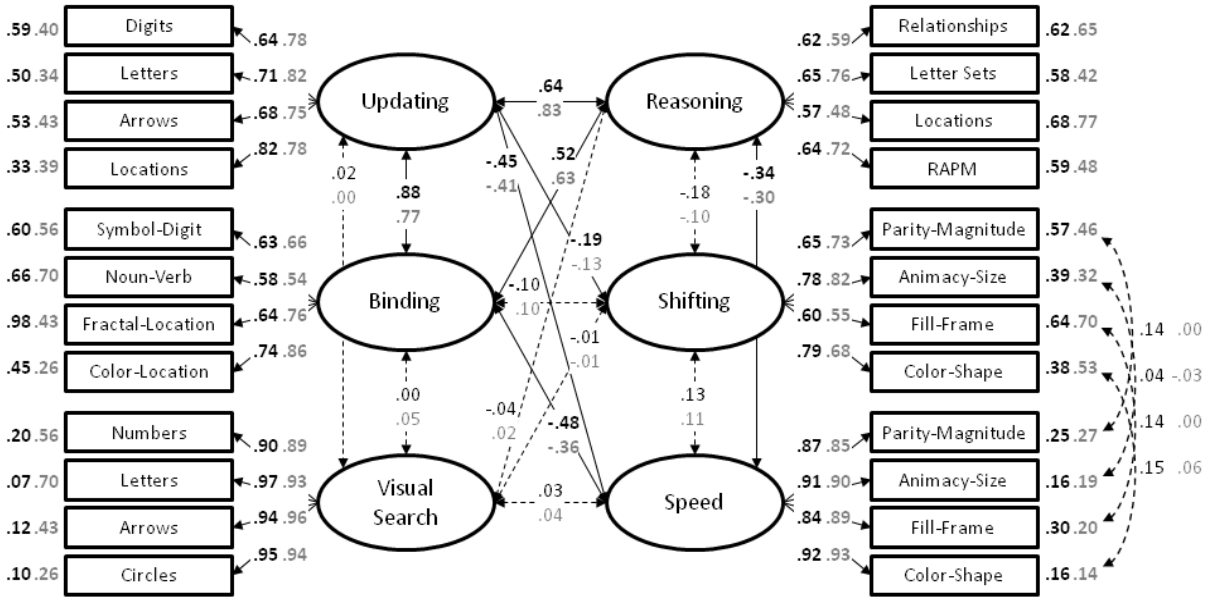

Figure 2. Measurement model for the theoretical transfer model for pretest data (printed in black) and posttest data (printed in gray). Rectangles represent manifest variables and ellipses latent factors. Single headed arrows represent linear regression and double-headed arrows correlations. Bold numbers and solid lines indicate significance ($p < .05$), and dotted lines non-significance. All latent factor variances were significant ($p < .05$). Manifest variables of the shifting and speed factor were correlated, because they originated from the same tasks.

2.4.2 Baseline Group Comparability

To evaluate whether baseline cognitive performance was comparable across groups, we ran Bayesian LME models on the level of constructs using the respective output measures as dependent variables, group as fixed effect, and participant and task as crossed-random effects. There was no evidence for group differences in baseline performance (all $BF_{H0} > 3$, see Table A3 in the Appendix) for any of the constructs except shifting, for which we found strong evidence in favor of a group difference with a $BF_{\mathrm{H1}} = 10.14$. To further investigate this group difference, we ran separate models comparing each group to the other groups. Group comparisons showed that the effect was primarily caused by a difference between the updating and binding groups (BF_{H1} = 91.56; updating vs. active control: BF_{H0} = 1.43; binding vs. active control: BF_{H0} = 1.54). Baseline comparisons on the task level are also listed in Table A3 in the Appendix.

2.4.3 Gains During Training

For each training task, we conducted a Bayesian LME model using the mean of the maximum load achieved during each training session as dependent variable, training session as fixed effect, and participant as random effect. We coded training session as linear contrast to evaluate monotonic trends rather than erratic fluctuations. As summarized in Table 5 and reflected by Figure 3, all training groups exhibited decisive evidence for an effect of session (all *BFs* > 100).

Table 5

Linear Contrasts of Training Effects on Performance During Training

Task	Estimate	95 % HDI	*BF*
Updating			
Digits	0.13	[0.12, 0.13]	$\mathbf{3.70*10^{278}}$
Letters	0.10	[0.09, 0.10]	$\mathbf{5.70*10^{242}}$
Arrows	0.06	[0.05, 0.06]	$\mathbf{2.00*10^{141}}$
Locations	0.08	[0.07, 0.08]	$\mathbf{7.80*10^{196}}$
Binding			
Symbol-Digit	0.08	[0.07, 0.08]	$\mathbf{3.20*10^{204}}$
Noun-Verb	0.09	[0.09, 0.10]	$\mathbf{8.60*10^{200}}$
Fractal-Location	0.13	[0.12, 0.13]	$\mathbf{1.97*10^{331}}$
Color-Location	0.11	[0.11, 0.12]	$\mathbf{6.90*10^{280}}$
Visual Search			
Numbers	2.17	[2.15, 2.19]	$\mathbf{4.50*10^{1039}}$
Letters	2.11	[2.09, 2.13]	$\mathbf{1.43*10^{999}}$
Arrows	2.16	[2.14, 2.16]	$\mathbf{1.86*10^{1050}}$
Circles	1.73	[1.71, 1.76]	$\mathbf{1.03*10^{727}}$

Note. Bold *BF* values indicate effects in favor of a training effect. Only participants with complete training data sets were included in the analyses. The dependent variable was maximum load achieved in each training session. The estimate is the mean of the sampling from the posterior distribution with 10000 iterations. HDI = highest density interval of the posterior distribution.

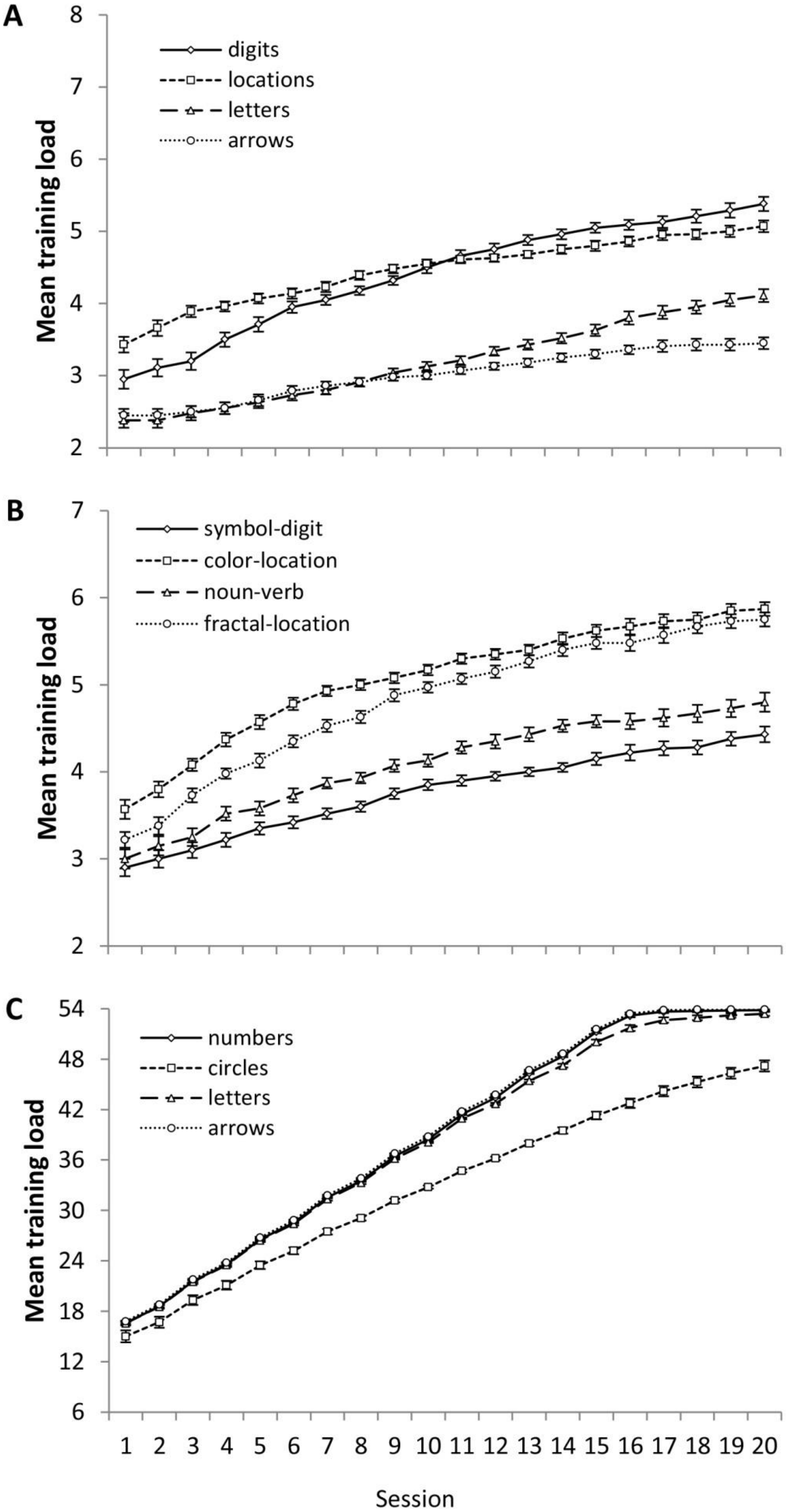

Figure 3. Performance gains during A) updating, B) binding, and C) active control training.

Error bars denote the 95% Cousineau-Morey confidence intervals for within-subjects comparisons (Baguley, 2012).

2.4.4 Motivation During Training

Figure 4 illustrates mean ratings of the three one-item training motivation measures (enjoyment, effort, and perceived fit between task difficulty and ability, cf. von Bastian & Eschen, 2016) over the course of the 20 training sessions. To examine whether motivation during training differed between the three training groups, we conducted Bayesian LME models separately for each of the measures as dependent variable, group and session (coded as linear contrast) as independent variables, and participant as random factor. For enjoyment experienced during training, we found strong evidence for the absence of a main effect of group (BF_{H0} = 29.03). Furthermore, enjoyment remained relatively stable throughout training, with strong evidence against a monotonic effect of session (BF_{H0} = 61.03). Moreover, there was decisive evidence against an interaction between group and the linear trend of session ($BF_{\mathrm{H0}} = 1.41*10^7$).

However, we found substantial evidence for group differences in effort spent during training (BF_{H1} = 3.77). Separate LMEs for each group comparison revealed anecdotal to strong evidence that participants of the active control group spent overall more effort during training in comparison to the updating group (BF_{H1} = 2.46) and the binding group (BF_{H1} = 11.76), with no differences between the binding and the updating group (BF_{H0} = 6.20). Decisive evidence for a linear effect of session ($BF_{H1} = 1.29*10^7$) reflected that effort spent on training decreased linearly over the course of the 20 training sessions. Moreover, there was decisive evidence for an interaction between group and the linear trend of session (BF_{H1} = 222.39), indicating different courses of effort during training for the three groups. Group comparisons revealed substantial to decisive evidence for an interaction of group and session between the updating and binding group (BF_{H1} = 6.47), and between the updating and active control group (BF_{H1} = 477.12). There was no evidence for an interaction between the binding and active control group (BF_{H0} = 636.82), indicating that these two groups did not differ in their effort across training sessions. Figure 4 suggests that the updating group showed more variation in effort during training than the other two groups. Whereas the other two groups started off with higher ratings that remained relatively stable (active control group) or steadily decreased (binding group), the updating group's rating of effort spent was first lower but then increased for a few sessions, followed by a decrease in effort until shortly before the end of the training, when effort ratings increased again).

Finally, there was anecdotal evidence for group differences in perceived fit between task difficulty and ability (BF_{H1} = 2.08), with the updating group's perceived fit between task difficulty and ability being overall worse than that of the active control group (BF_{H1} = 5.11) and the binding group (BF_{H1} = 2.21). The comparison of the binding and the active control group yielded substantial evidence for the absence of group differences in perceived fit (BF_{H0} = 7.43). Evidence decisively supported the absence of a linear trend of session (BF_{H0} = 246.12) as well as an interaction between group and the linear trend of session (BF_{H0} = 235.63).

In sum, the three training interventions were perceived to be equally enjoyable. However, whereas the active control group rated their effort spent as generally higher than the other two groups, the updating group varied more in their rating than the two other groups across the sessions. Also, the updating group perceived the fit between difficulty and ability overall as worse than the other two groups.

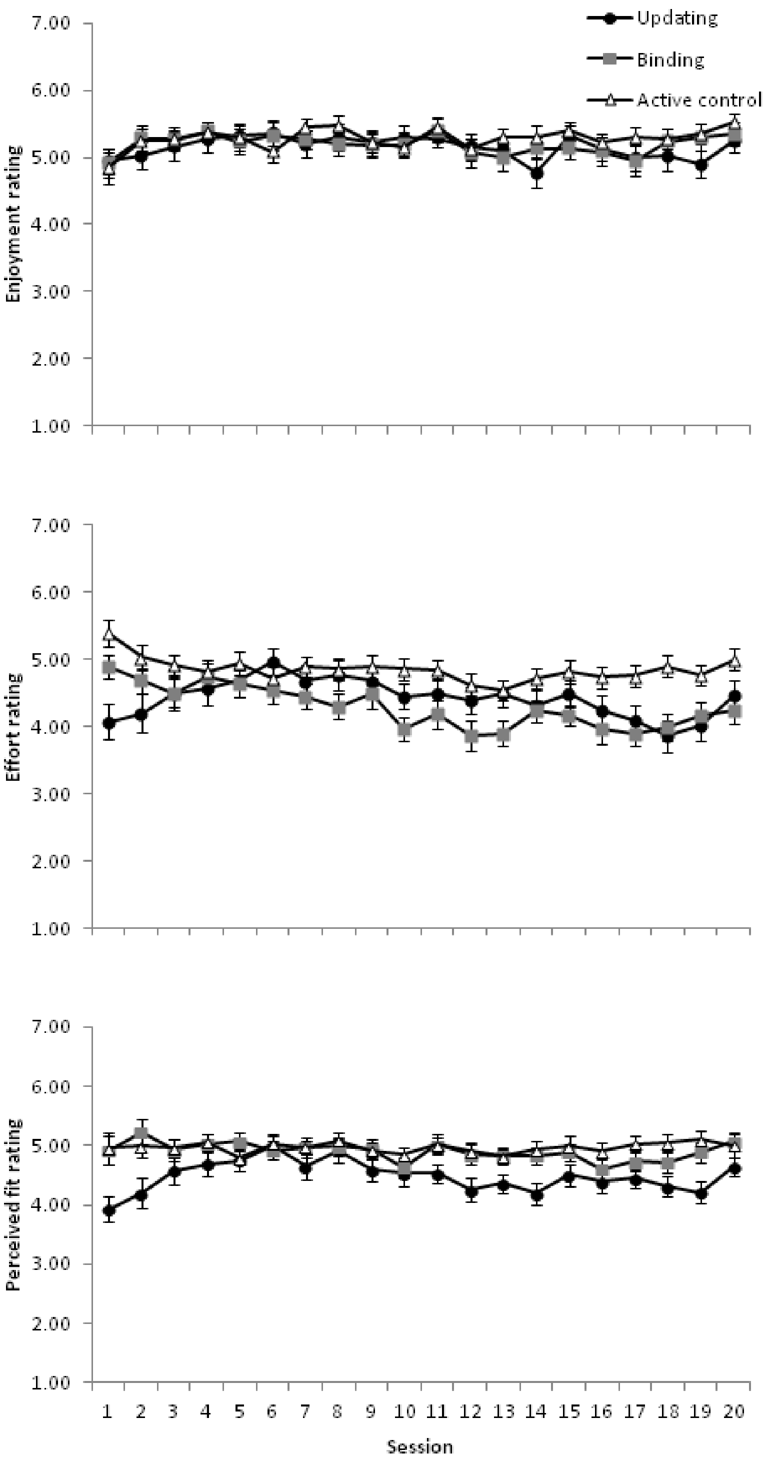

Figure 4. Motivation measures (i.e., enjoyment, effort and perceived fit) across the 20 training sessions. Error bars denote the 95% Cousineau-Morey confidence intervals for within-subjects comparisons (Baguley, 2012).

2.4.5 Gains from Pretest to Posttest

To evaluate effects of training on the practiced tasks, on specific working memory mechanisms, near transfer, and far transfer to cognitive abilities related to WM, we first computed standardized gain scores (i.e., mean of posttests scores minus mean of pretest scores divided by the pretest standard deviation; cf. von Bastian & Oberauer, 2013) for each participant and each task. We then conducted Bayesian LME models to estimate these gain scores on the level of constructs. Group served as fixed effect, and participant and task as crossed-random factors. Results of the Bayesian LME models are summarized in Table 6 and Figure 5.

Table 6

Parameter Estimates for Fixed Effects of the Bayesian Linear-Mixed Effects Models

		Grand Mean		Updating Group		Binding Group		Active Control Group	
Measure	*BF*	Estimate	95% HDI	Estimate	95% HDI	Estimate	95% HDI	Estimate	95% HDI
Training									
Updating	**71.67**	0.36	[-0.15, 0.86]	0.21	[0.11, 0.31]	-0.06	[-0.16, 0.04]	-0.15	[-0.24, -0.05]
Binding	**$7.08*10^4$**	0.35	[-0.44, 1.15]	-0.16	[-0.30, -0.02]	0.39	[0.26, 0.53]	-0.23	[-0.37, -0.10]
Visual search	**1.83**	0.05	[-0.28, 0.37]	-0.04	[-0.11, 0.02]	-0.06	[-0.12, 0.01]	0.10	[0.04, 0.16]
Mechanisms									
Focus switching	13.73	-0.03	[-1.11, 1.06]	0.09	[-0.04, 0.23]	0.02	[-0.11, 0.16]	-0.12	[-0.25, 0.02]
Removal	17.74	0.03	[-0.79, 0.87]	-0.02	[-0.17, 0.13]	-0.09	[-0.24, 0.05]	0.11	[-0.03, 0.26]
Interference resolution	45.01	0.32	[-0.38, 1.01]	-0.06	[-0.18, 0.07]	0.00	[-0.13, 0.12]	0.06	[-0.06, 0.18]
Far transfer									
Reasoning	74.92	0.23	[-0.32, 0.77]	-0.03	[-0.12, 0.06]	0.02	[-0.07, 0.11]	0.01	[-0.08, 0.09]
Shifting	47.86	-0.05	[-0.70, 0.59]	0.04	[-0.09, 0.16]	-0.06	[-0.18, 0.06]	0.02	[-0.10, 0.14]
Speed	3.15	0.54	[0.14, 0.93]	0.07	[-0.02, 0.15]	0.03	[-0.06, 0.11]	-0.09	[-0.18, -0.01]
Inhibition	35.62	0.10	[-0.71, 0.93]	-0.06	[-0.19, 0.07]	0.09	[-0.03, 0.21]	-0.03	[-0.15, 0.09]

Note. Speed-based output variables (i.e., visual search, removal speed, focus switching, shifting, and speed) were coded so that higher estimates reflect better performance at posttest. Bold *BF* values indicate evidence in favor of group differences. The estimate is the mean of the posterior distribution with 10000 iterations. HDI = highest density interval of the posterior distribution.

Criterion tasks. There was substantial to decisive evidence for group differences in gain scores in the updating (BF_{H1} = 71.67) and in the binding (BF_{H1} = 7.08*10^4) tasks, and anecdotal evidence for group differences in the visual search tasks (BF_{H1} = 1.83). We followed up on this analysis to assess which groups differed from each other, running the same LME for the comparisons of the experimental groups with the active control group. For the updating tasks, we found substantial to very strong evidence for higher gain scores in the updating in comparison to the active control group (BF_{H1} = 144.25) and the binding group (BF_{H1} = 4.81). The comparison of the binding and the active control group yielded substantial evidence for a null effect (BF_{H0} = 3.32). Group comparisons for the binding tasks revealed evidence for a benefit of the binding group in comparison to the active control (BF_{H1} = 2.09*10^4) and the updating group (BF_{H1} = 239.31). However, there was substantial evidence for the absence of gain score differences between the active control and the updating group (BF_{H0} = 8.28). For the visual search tasks, we found substantial evidence for higher gain scores (i.e., smaller residuals) in the active control group in comparison to the binding group (BF_{H1} = 3.32), with the evidence being more ambiguous in comparison to the updating group (BF_{H1} = 2.25). The comparison of the two experimental groups yielded strong evidence for the absence of differences in gain scores (BF_{H0} = 10.94). To summarize, each training group showed evidence for improvements in the tasks they had practiced for five weeks, but neither experimental group showed evidence for near transfer to the respective other working memory tasks. Participants of the updating group yielded comparable gain scores in the binding tasks as those of the active control group and participants of the binding group did not differ in the updating tasks when compared with participants of the active control group.

Mechanisms of WM measures. Next, we investigated group differences in gain scores for focus switching, removal, and interference resolution. We found strong evidence for the absence of group differences in focus switching (BF_{H0} = 13.73), and (BF_{H0} = 17.74), and interference resolution (i.e., intrusion probes, BF_{H0} = 45.01). To identify the source of the binding group's strong practice effects in the binding, we additionally explored whether there was any evidence supporting group differences in gain scores for matching (i.e., positive) or distractor (i.e., negative) probes (see Table 7). We found decisive evidence for group differences in gain scores for matches (BF_{H1} = $6.40*10^5$), with the binding group exhibiting higher gain scores (M = 0.42 [0.27, 0.56]) than the active control (M = -0.32 [-0.46, -0.18], BF_{H1} = $2.41*10^5$) and the updating group (M = -0.10 [-0.24, 0.05], BF_{H1} = 141.68). In contrast, evidence was ambiguous for the comparison between the updating and the active control group (BF_{H0} = 1.80). Evidence regarding group differences in gain scores for distractor probes was ambiguous (BF_{H0} = 1.15), with the evidence slightly favoring a difference between the binding (M = 0.17 [0.05, 0.28]) and active control (M = -0.08 [-0.20, 0.03], BF_{H1} = 2.40) and the updating group (M = -0.08 [-0.20, 0.03], BF_{H1} = 1.39), whereas evidence strongly supported the absence of a difference between the updating and the active control group (BF_{H0} = 11.15). To summarize, there was no evidence of training-specific change for any of the proposed mechanisms. Instead, the ability to recognize matches was improved in the binding group in comparison to the active control and updating group.

Table 7

Binding Task Probes Performance as a Function of Training Group and Time of Assessment

	Group							
	Updating		Binding		Active Control		Reliabilities	
Probes	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Matches								
Symbol-digit	.72 ± .10	.69 ± .10	.71 ± .10	.73 ± .13	.71 ± .11	.67 ± .14	.72	.81
Noun-verb	.78 ± .11	.75 ± .12	.78 ± .10	.76 ± .13	.76 ± .10	.72 ± .13	.77	.83
Fractal-location	.72 ± .09	.72 ± .11	.71 ± .12	.84 ± .09	.70 ± .11	.68 ± .11	.73	.83
Color-location	.74 ± .10	.77 ± .10	.74 ± .12	.82 ± .09	.72 ± .11	.72 ± .10	.76	.76
Distractors								
Symbol-digit	.73 ± .12	.75 ± .13	.73 ± .14	.76 ± .14	.74 ± .13	.77 ± .13	.71	.73
Noun-verb	.85 ± .11	.85 ± .12	.84 ± .14	.86 ± .14	.83 ± .12	.84 ± .12	.74	.78
Fractal-location	.67 ± .12	.72 ± .14	.70 ± .13	.81 ± .13	.70 ± .13	.72 ± .13	.66	.75
Color-location	.79 ± .11	.83 ± .13	.76 ± .13	.83 ± .12	.79 ± .11	.82 ± .13	.67	.74

Note. Values denote *M*s ± *SD*s.

Far transfer gains. Next, we examined whether training led to transfer to abilities related to updating and binding. Neither experimental group showed evidence for any transfer effects (all $BF_{H0} > 3.15$), although abilities correlated at least moderately at pretest. For the sake of completeness, we also conducted an LME model for inhibition, even though these tasks had to be excluded from the theoretical transfer model. There was very strong evidence for the absence of group differences in inhibition ($BF_{H1} = 35.62$).

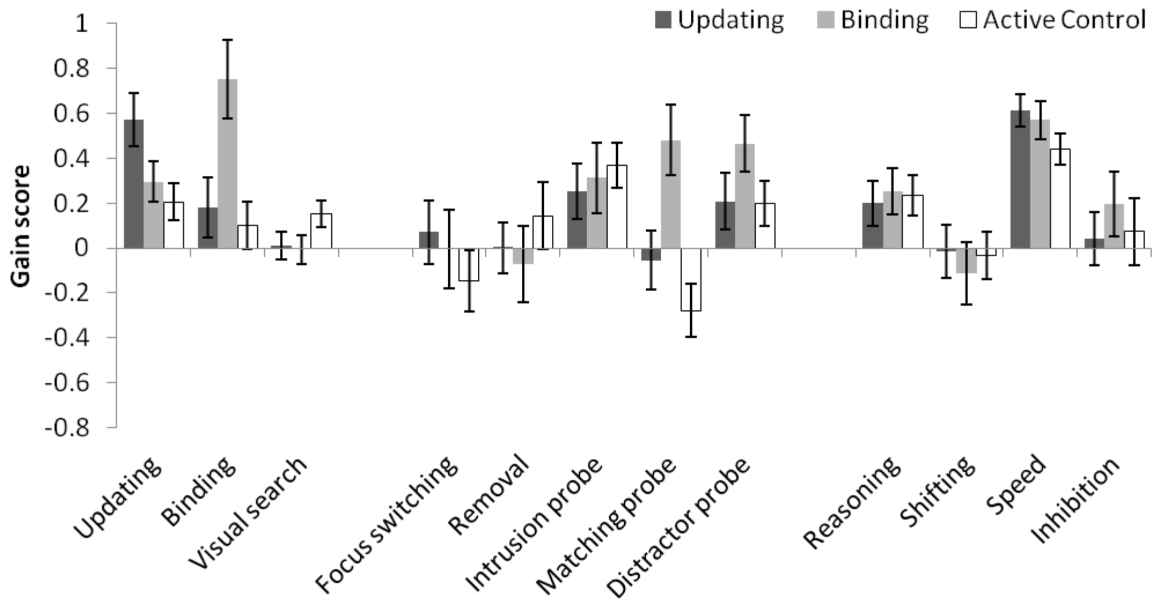

Figure 5. Gain scores on the level of constructs for training, WM mechanisms and far transfer. Note that any speed-based output variable (i.e., visual search, removal speed, focus switching, shifting, and speed) was reversely coded, hence higher gain scores reflect better performance at posttest. Error bars denote 95 % confidence intervals.

2.4.6 Strategy Use

Frequency of strategy use and average ratings of perceived helpfulness are reported in Table 8. To investigate whether the frequency of strategy use differed between the training groups, we conducted a Pearson chi-square test, which revealed strong evidence for group differences (BF_{H1} = 36.96). Subsequent group comparisons showed that strategy use was more frequent in the updating and binding group in comparison to the active control group (BF_{H1} = 4.70 and BF_{H1} = 123.60, respectively). The two experimental groups did not differ in the frequency of strategy use (BF_{H0} = 4.04). Next, we examined group differences in the ratings of perceived helpfulness with a Bayesian ANOVA. There was strong evidence for group differences (BF_{H1} = 24.61). Whereas the updating and binding group rated their strategies equally helpful (BF_{H0} = 4.23), the active control group rated the helpfulness of their strategies worse than the updating (BF_{H1} = 5.39) and worse than the binding group (BF_{H1} = 64.67). In sum, the two experimental groups reported to have relied more on strategies and rated their helpfulness higher than the active control group.

Table 8

Retrospective Report of Strategy Use

Group	Strategy use	Helpfulness
Updating	84 %	4.18 ± 0.93
Binding	91 %	4.28 ± 0.77
Active Control	65 %	3.65 ± 0.97

Note. Percentage of participants stating that they used a strategy. The perceived helpfulness scale ranged from 1 (*not at all*) to 5 (*very*).

2.5 Discussion

In this study, we pursued three goals. First, we investigated whether five weeks of practice enhances updating and binding performance. Second, we assessed the specific WM mechanisms that were potentially improved through updating and binding training. Third, we examined transfer of updating and binding training to untrained, structurally different WM tasks (i.e., near transfer), and to closely related abilities (i.e., far transfer to reasoning, processing speed, shifting, and inhibition).

Concerning our first goal, we conducted linear trend analyses that revealed decisive evidence for monotonic performance increases in each training group and task across the 20 training sessions. We additionally measured training gains with criterion tasks that were administered at pretest and posttest to assess training effects against baseline performance. We found decisive evidence for superior performance of the updating and binding group in their respective training tasks. The active control group practicing visual search also showed better performance in their training tasks than the binding group, but the data were inconclusive when compared to the updating group. Possibly, the visual search criterion tasks with a maximum set size of 11 were too easy and, hence, insensitive to detect differences between the groups.

Next, we analyzed training gains in focus switching, removal, and interference resolution. Contrary to our expectations, neither WM training group showed improvements in any of the proposed WM mechanisms over and above those of the active control. Finally, we investigated near and far transfer. In line with previous studies (e.g., Schmiedek, Hildebrandt, Lövdén, Lindenberger, & Wilhelm, 2009; Wilhelm et al., 2013) performance in updating and binding tasks was strongly correlated in our study, with the latent factors sharing 77 % (pretest) and 41 % (posttest) of variance. However, contrary to our expectations, we did not find evidence for any transfer from updating to binding or vice versa. Also, even though the transfer model confirmed moderate correlations between working memory and reasoning (Kyllonen & Christal, 1990), processing speed (Schmiedek et al., 2007), shifting, and inhibition (Miyake et al., 2000), we found no evidence for any far transfer effects.

2.5.1 Strengths and Limitations

One of the major strengths of our study is the theory-driven selection of training and transfer tasks. We based our study on the three-embedded-components model of working memory (Oberauer, 2009; Oberauer & Hein, 2012), which describes WM as three different levels of information selection, each level narrowing down the information content of the preceding one. A theory-based approach of task selection allows for determining whether an improvement in a cognitive task is accomplished through an increase in WM capacity or enhanced WM efficiency (cf. von Bastian & Oberauer, 2014). Moreover, we systematically investigated specific WM mechanisms that have been discussed as potential mediators of near and far transfer effects.

Another strength of our study was the large sample size with 59 to 72 participants per group. According to the meta-analysis by Melby-Lervåg et al. (2016) that included 87 cognitive training studies, most previous studies compared groups of 30 participants or less, with only one study with group sizes larger (i.e., Estrada, Ferrer, Abad, Román, & Colom, 2015 with 114 to 193 participants per group) and one other study with group sizes comparable (i.e., Sprenger et al., 2013 with 57 and 70 participants per group in experiment 1) to ours. Indeed, the resulting Bayesian evidence was at least substantial in our study.

A further strength of our study is the inclusion of an active control group. Although meta-analyses sometimes show no effect of the type of control group (i.e., active or passive) on the average transfer effect (e.g., Au et al., 2015; Karbach & Verhaeghen, 2015; but see Dougherty et al., 2015), including an active control group is still important from a methodological standpoint, as the absence of a statistical difference cannot exclude confounds from non-specific intervention effects in any given single study. An optimal control group should demand only little WM, and participants should perceive the control intervention as a believable and potentially effective intervention (cf. Morrison & Chein, 2011; Shipstead, Hicks, & Engle, 2012; von Bastian & Oberauer, 2014). Hence, the control treatment should be as similar as possible to the experimental treatment. We argue that visual search training met these criteria well. First, visual search did not correlate with the other cognitive abilities assessed as demonstrated by the transfer model (Figure 2). Second, training conditions were identical across interventions (e.g., number and duration of training sessions, adaptivity, and the amount of feedback and experimenter contact). Moreover, motivational measures assessed during training showed that participants of the control group perceived training as equally enjoyable as the two experimental groups. The only substantial difference we observed occurred for effort, with participants of the active control group however reporting having spent more effort during training than the experimental groups. We can only speculate about reasons for this difference, but possibly, ratings of the effort spent are a consequence of the perceived performance in that session. More specifically, as the active control group showed steeper increases in training performance than the other two groups, they on average experienced more level-ups per session than the two experimental groups. Hence, participants might have used their progress in a session in terms of level-ups as reference to retrospectively rate their effort in this session.

Finally, we measured each cognitive ability with four tasks. The application of multiple indicators to assess a construct minimizes the problems arising from task impurity (Miyake & Friedman, 2012). More specifically, the score of a task reflects both cognitive ability as well as systematic and random influences (e.g., Shipstead, Redick, et al., 2012). For example, in our study, the latent updating factor explained between 41 % and 67 % of variance in the single updating tasks. Hence, there was still a considerable amount of variance from task-specific sources. Therefore, transfer effects found for single tasks may reflect changes in the task-specific portions of variance rather than an enhancement of the underlying ability (Schmiedek et al., 2010). To extract the ability-specific effects, multiple measures of an ability are needed. An often-voiced concern about measuring each ability with four indicators is that it increases testing time. Green et al. (2014) argued that such increased testing time could lead to ego-depletion or fatigue effects and, hence, obscure training and transfer gains. However, in a recent reanalysis of data from cognitive test batteries that comprised at least 10 tasks and lasted not less than two hours, we found no evidence for ego-depletion (cf. De Simoni, Luethi, Oberauer, & von Bastian, 2017).

A potential drawback of our study is that training was self-administered at home. Despite regularly keeping in touch with participants during training and closely monitoring their training progress, experimental control was less than if participants had completed the training in the laboratory. Laboratory-administered interventions have been argued to strengthen the commitment and motivation of participants (cf. Lampit, Hallock, & Valenzuela, 2014). However, regular personal contact with the experimenter also yields an increased risk of experimenter effects (e.g., participants act as expected from the experimenter, cf. Rosenthal, et al. 2005), possibly confounding training and transfer data. Moreover, although laboratory-based training interventions increase control over the environment during training (e.g., minimized impact of distractions), home-administered training interventions increase ecological validity of the results. Empirically, it is yet unclear whether WM training effectiveness is affected by whether it training is administered at home or in the laboratory. However, although some meta-analyses that include different age groups support an effect (e.g., Lampit et al., 2014; Schwaighofer et al., 2015), others that focus solely on young adults do not (e.g., Au et al., 2015). Indeed, several laboratory-based cognitive training studies did not find any transfer effects either (e.g., Linares et al., 2017; Redick et al., 2013), whereas other self-administered interventions were successful in establishing even far transfer (e.g., Jaeggi, Buschkuehl, Shah, & Jonides, 2014; von Bastian & Oberauer, 2013).

Taken together, it seems unlikely that the self-administered training regime is responsible for the lack of transfer.

Finally, one could argue that young adults are simply not an adequate target for training interventions, as they are at the peak of their cognitive functioning, and, hence, room for additional improvement is small at best (e.g., Bherer et al., 2008; Karbach & Kray, 2009, see also Titz & Karbach, 2014 for a review). However, in a recent training study with older adults that followed a similar design and yielded comparably strong Bayesian evidence as the present study, we also found no transfer effects (Guye & von Bastian, 2017). Moreover, one could argue that individual differences such as personality, motivation, and beliefs might have obscured potential transfer effects (cf. Jaeggi, Buschkuehl, Shah, & Jonides, 2014). Therefore, we investigated the impact of individual differences on training trajectories across the two studies, but found little evidence for a potential role of individual differences (Guye et al., 2017). Therefore, it seems unlikely that age or other individual differences can explain the absence of transfer in the present study.

2.5.2 Implications and Future Directions

Participants of the updating and binding group improved in the tasks they practiced, however there was no evidence for an improvement in WM mechanisms beyond that of the active control group. Moreover, there was substantial to very strong evidence supporting the absence of any near and far transfer effects. Thus, neither WM capacity nor efficiency of the proposed WM mechanisms were increased during training. Hence, the question remains, what did improve during training? Our data suggest both task-specific and material-specific changes in information processing. First, the absence of any transfer effects despite strong practice effects suggests that participants in the WM training groups acquired task-specific strategies (cf. Soveri et al., 2017). Indeed, 84 % of the participants in the updating group and

91 % of the participants in the binding group reported having used strategies they experienced as helpful for completing the training tasks. Qualitative reports of the specific strategies used suggest that these strategies often depended on the task material (e.g., inventing a story around verbal memoranda, or memorizing arrows in the spatial task as watch hands).

Second and possibly related to the use of material-specific strategies, we found evidence for enhanced familiarity processing. More specifically, the binding group showed strongly enhanced performance in the practiced tasks, which however was not specific to the resolution of interference (i.e., intrusion probes). Hence, changes in performance were due to a better hit rate, that is, correct recognition of matching (i.e., positive) probes and rejection of distractor probes. Indeed, in additional explorative analyses we found decisive evidence for better performance on matching probes and a tendency toward improved performance on distractor probes. Examining performance on the different probes allows for distinguishing between two recognition processes: familiarity (i.e., the memory of an item) and recollection (i.e., the memory of the binding between an item and its context, cf. Oberauer, 2008). In matching and distractor probes, familiarity and recollection signals drive recognition decisions toward the same response (i.e., toward a YES response in positive, and a NO response in distractor probes). However, intrusion probes generate conflict, as signals of familiarity and recollection go in opposite directions (toward a YES response based on familiarity, and to a NO response based on recollection, cf. Oberauer, 2008). A selective improvement in hit rates would suggest that the binding group simply shifted their response bias toward giving more YES responses. However, whereas the binding group outperformed the active control in matching and distractor probes, they similarly in intrusion probes. Hence, rather causing a mere shift in response bias, binding practice improved primarily familiarity and, to a lesser degree, recollection. In contrast, the active control group improved in intrusion probes (cf. Figure 4), but also showed poorer performance in matching probes after training, a pattern in line with a shift in response bias towards NO responses in the active control group in comparison to the binding group. Taken together, the binding group showed superior item memory in contrast to the two other groups, this ability, however, seems to be highly task-specific, as no transfer to other abilities was observed.

The absence of transfer effects alongside task- and material-specific effects after intensive practice is in line with decades of research in skill acquisition, which showed that expertise (e.g., in digit memory span, cf. Ericsson, Chase, & Faloon, 1980) is highly specific and very limited in terms of transfer to new tasks (cf. Ericsson et al., 1980; Healy, Wohldmann, Sutton, & Bourne, 2006; Lewandowsky & Thomas, 2009). For example, Healy et al. (2006) showed that participants who learned to use a computer mouse that reverses either vertical, horizontal or a combination of both movements, could not transfer this ability to one of the other reversal conditions (e.g., transfer from vertical to horizontal movements or vice versa).

2.6 Conclusion

To conclude, this study reports strong evidence for the absence of transfer effects after WM training. Although we found large practice effects and moderate to strong correlations between the cognitive abilities at pre- and posttest, our data consistently supported the absence of near and far transfer. The absence of any transfer effects suggests that neither updating nor binding training enhanced WM capacity or efficiency, but instead encouraged the acquisition of task-specific strategies. Indeed, participants reported to strongly rely on strategies during training. Thus, the current findings suggest that WM is a largely stable trait that cannot be fundamentally changed through training interventions.

2.7 Appendix

Table A1

Excluded Data from Pretest and Posttest

Tasks	**Pretest**	**Posttest**
Technical Issues		
Updating letters	1	
Removal arrows	1	
Compliance		
Updating arrows	1	
Binding symbol-digit		1
Binding fractal-location	1	2
Binding color-location	2	2
Visual search letters	8	8
Visual search circles	1	1
Removal letters	1	
Removal arrows	4	
Removal locations		1
Reasoning relationships	1	1
Reasoning letter sets		1
Reasoning locations	1	3
Reasoning RAPM		2
Shifting fill-frame	1	
Inhibition numerical Stroop	1	
Inhibition global-local	1	
Total	**25**	**22**

Table A2

Correlations between Tasks at Pretest and Posttest

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1. Digits (upd)		**.71**	**.58**	**.53**	**.41**	**.34**	**.39**	**.49**	-.01	.00	.04	.00	**.44**	**.63**	**.30**	**.43**	-.09	-.14	.00	**-.22**	**-.24**	**-.30**	**-.29**	**-.31**	-.07	**.27**	.05	-.06
2. Letters (upd)	**.58**		**.62**	**.66**	**.36**	**.30**	**.33**	**.49**	-.02	.00	.03	-.05	**.42**	**.52**	**.34**	**.47**	-.05	**-.18**	-.08	**-.30**	**-.31**	**-.32**	**-.33**	**-.38**	-.03	**.26**	.05	.01
3. Arrows (upd)	**.49**	**.57**		**.63**	**.45**	**.41**	**.39**	**.54**	.00	.01	.04	-.03	**.37**	**.47**	**.35**	**.39**	-.01	-.06	.02	**-.26**	**-.31**	**-.31**	**-.33**	**-.36**	-.04	**.24**	.02	.08
4. Locations (upd)	**.49**	**.58**	**.56**		**.47**	**.43**	**.50**	**.67**	-.01	.01	.01	-.01	**.31**	**.44**	**.36**	**.44**	.05	-.09	.05	**-.15**	**-.24**	**-.30**	**-.28**	**-.30**	.12	**.26**	.08	.11
5. Symbol-Digit	**.32**	**.35**	**.34**	**.45**		**.55**	**.55**	**.50**	-.01	-.01	.03	.00	**.17**	**.38**	.12	**.31**	.06	-.01	-.07	-.13	**-.18**	**-.33**	**-.24**	**-.26**	-.07	**.20**	.02	-.06
6. Noun-Verb	**.44**	**.33**	**.40**	**.43**	**.46**		**.42**	**.40**	.00	.00	.07	.03	**.17**	**.22**	**.26**	**.31**	.05	-.04	**-.16**	-.07	**-.18**	**-.22**	**-.17**	**-.19**	-.09	**.21**	-.01	-.04
7. Fractal-Location	**.34**	**.40**	**.30**	**.56**	**.41**	**.43**		**.67**	.06	.07	.07	.04	**.25**	**.42**	**.25**	**.31**	.07	.02	-.01	-.02	**-.28**	**-.32**	**-.25**	**-.22**	.03	**.25**	.06	.02
8. Color-Location	**.45**	**.43**	**.42**	**.67**	**.40**	**.32**	**.52**		-.01	.05	.05	.02	**.24**	**.43**	**.27**	**.37**	.12	.06	.10	-.05	**-.24**	**-.24**	**-.24**	**-.21**	.07	**.28**	.08	.11
9. Digits (vs)	.01	-.01	.03	-.01	.00	-.04	-.01	-.02		**.83**	**.85**	**.83**	.02	.06	.00	.07	.01	.05	.03	.01	.00	-.01	.01	.02	-.02	.02	.02	-.02
1. Letters (vs)	-.05	-.02	.02	.02	-.07	-.05	-.01	.02	**.85**		**.89**	**.87**	.01	.04	-.06	.03	.00	.04	.05	-.02	.03	-.02	-.01	.00	.02	-.02	.05	-.06
11. Arrows (vs)	.00	.01	.03	-.01	.01	.0.	-.02	.01	**.87**	**.90**		**.89**	.07	.04	.02	.05	.00	.02	.01	-.01	.00	-.01	-.02	-.01	-.01	-.03	.04	-.06
12. Circles (vs)	-.02	-.02	.02	.01	-.03	-.01	-.02	.03	**.84**	**.92**	**.88**		.00	-.02	-.03	.01	.01	.05	.04	.02	.01	.02	.05	.02	.01	-.07	.04	-.05
13. Relationships	**.30**	**.30**	**.25**	**.22**	.08	**.16**	**.15**	**.20**	-.04	-.03	-.01	.00		**.54**	**.31**	**.44**	-.10	-.09	-.04	**-.19**	-.14	**-.16**	-.12	**-.16**	-.08	.10	**.21**	-.03
14. Letter Sets	**.44**	**.39**	**.31**	**.33**	**.22**	**.27**	**.29**	**.28**	.02	-.02	.04	.00	**.45**		**.26**	**.57**	-.14	-.13	.03	**-.19**	**-.20**	**-.24**	**-.24**	**-.25**	-.03	**.22**	.14	-.06
15. Locations	**.19**	**.18**	**.22**	**.39**	**.22**	**.27**	**.31**	**.24**	-.08	-.05	-.04	-.02	**.44**	**.36**		**.42**	.01	-.04	.00	**-.18**	**-.23**	**-.20**	**-.32**	**-.20**	-.01	**.22**	**.18**	.06
16. RAPM	**.41**	**.35**	**.36**	**.38**	**.26**	**.23**	**.25**	**.30**	-.03	-.07	-.01	-.05	**.38**	**.46**	**.29**		.00	.00	.02	**-.17**	**-.14**	-.10	-.12	-.09	-.02	**.24**	.13	.05
17. Parity-Magnitude (sc)	**-.15**	**-.17**	**-.25**	-.09	.07	-.04	-.05	-.06	.06	-.01	.00	.04	**-.17**	**-.22**	-.14	-.13		**.59**	**.33**	**.53**	-.05	.04	.09	.06	.03	-.02	**-.15**	.08
18. Animacy-Size (sc)	.00	-.07	**-.17**	-.09	-.04	-.13	-.04	-.10	.03	-.02	-.05	.00	-.06	-.10	-.08	-.07	**.52**		**.49**	**.54**	.02	.10	**.17**	.10	.05	**-.16**	**-.19**	-.04
19. Fill-Frame (sc)	-.05	.01	-.03	.01	-.07	-.05	.03	.07	.05	.00	-.03	.03	-.07	-.08	-.07	-.05	**.36**	**.51**		**.38**	-.03	.08	.05	.00	.09	-.03	**-.16**	.09
2. Color-Shape (sc)	-.13	**-.21**	**-.22**	**-.15**	-.03	-.06	-.10	**-.17**	.02	-.05	-.02	-.03	-.10	**-.16**	-.07	-.11	**.51**	**.60**	**.48**		.12	**.17**	**.22**	**.19**	.09	-.10	**-.22**	-.03
21. Parity-Magnitude (ps)	**-.38**	**-.34**	**-.37**	**-.35**	**-.27**	**-.26**	**-.26**	**-.39**	.01	.02	.03	.03	**-.30**	**-.25**	**-.34**	**-.26**	**.16**	**.14**	-.01	.13		**.71**	**.71**	**.76**	**.16**	**-.45**	-.05	-.13
22. Animacy-Size (ps)	**-.29**	**-.32**	**-.31**	**-.33**	**-.32**	**-.22**	**-.30**	**-.32**	.00	.01	.02	.01	**-.17**	**-.16**	**-.23**	**-.18**	.06	.14	-.05	.09	**.77**		**.82**	**.83**	.13	**-.34**	-.03	-.06
23. Fill-Frame (ps)	**-.30**	**-.28**	**-.36**	**-.37**	**-.29**	**-.26**	**-.19**	**-.37**	.02	.02	.01	.02	**-.21**	**-.29**	**-.31**	**-.26**	**.16**	**.22**	.07	**.20**	**.72**	**.77**		**.84**	.08	**-.39**	-.12	-.10
24. Color-Shape (ps)	**-.38**	**-.35**	**-.34**	**-.35**	**-.31**	**-.27**	**-.25**	**-.36**	.02	.02	.03	.02	**-.20**	**-.25**	**-.30**	**-.15**	.12	**.16**	-.06	**.16**	**.78**	**.84**	**.78**		.09	**-.39**	-.02	-.03
25. Numerical Stroop	-.08	-.06	.07	.00	-.01	-.05	-.07	-.01	.04	.02	-.02	.01	-.13	-.04	.09	-.03	.00	-.03	.01	-.01	.14	.13	.03	.09		.01	-.04	.07
26. Color Stroop	**.21**	**.25**	**.28**	**.37**	**.19**	**.22**	**.28**	**.31**	.01	.00	-.06	.01	.12	.06	**.24**	.13	-.02	-.04	.05	-.05	**-.44**	**-.43**	**-.37**	**-.47**	.08		-.08	.13
27. Global-Local	.03	.13	.10	.04	-.02	.06	.04	.11	-.09	-.01	-.02	-.06	**.28**	.14	**.17**	.14	-.09	**-.15**	-.13	**-.16**	-.02	.03	-.06	.02	.05	.03		-.14
28. Simon	.11	.06	**.14**	**.31**	.05	.08	**.16**	**.23**	-.04	-.02	-.09	.03	.08	-.02	**.18**	.01	.05	-.01	.00	-.05	**-.17**	**-.22**	**-.21**	**-.17**	-.03	.12	-.03	

Note. Correlations at pretest are illustrated below the diagonal and correlations at posttest above. Upd = Updating; vs = Visual search; sc = switching costs; ps = processing speed.

Table A3

Baseline Differences at Pretest on the Levels of Construct and Task

	Bayes Factor			
Measure	Overall	UPD vs. AC	BIN vs. AC	UPD vs. BIN
Updating	3.04	3.04	3.88	**1.25**
Digits	3.83	4.78	4.99	4.09
Letters	**1.67**	4.34	**1.16**	**2.43**
Arrows	1.83	4.14	2.69	1.27
Locations	3.08	2.98	5.22	3.18
Binding	24.30	5.86	5.95	7.11
Symbol-Digit	3.24	4.14	4.82	3.78
Verb-Noun	3.54	4.77	3.51	4.01
Fractal-Location	3.31	4.61	4.41	5.00
Color-Location	3.99	4.68	4.66	3.81
Visual search	5.90	3.54	2.92	2.00
Numbers	4.08	5.10	4.90	4.96
Letters	4.03	4.82	5.12	4.98
Arrows	4.27	5.15	5.21	5.02
Locations	4.18	5.12	5.18	4.93
Focus switching	43.74	8.68	6.64	10.27
Digits	3.25	2.96	5.18	3.68
Letters	4.18	5.13	5.09	4.85
Arrows	2.14	2.38	1.96	4.86
Locations	3.89	4.92	4.97	4.22
Removal	36.98	7.29	5.99	9.15
Digits	3.02	3.25	3.07	5.00
Letters	1.79	2.04	4.87	1.45
Arrows	**1.04**	1.78	**1.16**	2.65
Locations	4.06	5.07	4.74	4.86
Interference Resolution	45.92	8.68	8.46	7.89
Symbol-Digit	2.15	3.33	3.98	1.58
Verb-Noun	3.60	5.11	3.98	4.21
Fractal-Location	1.72	1.08	4.49	2.58
Color-Location	3.73	5.10	4.55	4.12
Other Probe Conditions				
Matches	19.84	4.07	4.90	6.86
Symbol-Digit	4.16	5.10	5.08	5.03
Verb-Noun	3.81	5.04	4.22	4.69
Fractal-Location	2.51	2.45	2.54	5.02
Color-Location	2.13	4.75	1.58	2.69
Distractor	30.06	7.71	6.48	6.04
Symbol-Digit	3.24	3.58	3.35	4.99
Verb-Noun	3.84	4.00	4.57	4.99
Fractal-Location	2.91	2.63	4.96	3.53
Color-Location	2.04	5.06	1.73	2.39
Reasoning	12.66	6.09	4.60	2.81
Relationships	1.33	5.07	**1.11**	1.51
Letter sets	3.45	4.87	3.53	4.26

Locations	1.55	3.43	3.02	**1.07**
RAPM	3.33	4.54	3.04	4.57
Shifting	**10.14**	1.43	1.54	**91.56**
Parity-Magnitude	**1.32**	2.27	2.89	**4.81**
Animacy-Size	**4.40**	1.52	1.95	**20.34**
Fill-Frame	**1.18**	4.77	**1.10**	**2.80**
Color Shape	**2.77**	1.25	2.76	**6.55**
Processing Speed	4.65	3.84	1.91	2.53
Parity-Magnitude	2.76	4.60	2.76	3.71
Animacy-Size	2.17	2.72	1.99	4.45
Fill-Frame	2.34	5.10	2.39	2.35
Color Shape	1.36	1.89	1.07	4.32
Inhibition	3.36	2.00	8.85	**1.92**
Numerical Stroop	1.86	**1.61**	4.21	3.22
Color Stroop	3.86	5.12	4.43	4.53
Global-local	**1.79**	3.03	1.39	**1.91**
Simon	2.77	2.25	4.08	4.27

Note. Bold values indicate *BF*s in favor of the H_1 (i.e., baseline differences). We conducted Bayesian liner mixed-effects models to evaluate baseline differences on the level of construct, and Bayesian ANOVA and Bayesian two-sided t-tests to calculate baseline differences on the level of task. UPD = updating group; AC = active control group; BIN = binding group.

3 No Consistent Evidence for Ego-Depletion Effects across Multiple Cognitive Tasks and Domains

Carla De Simoni[1], Matthias S. Luethi[2], Klaus Oberauer[3], and Claudia C. von Bastian[4]

[1] University of Zurich, Department of Psychology

[2] The Hebrew University of Jerusalem, Department of Psychology

[3] University of Zurich, Department of Psychology

[4] Bournemouth University, Department of Psychology

Submission status

Submitted for publication.

Authors' contributions

All authors contributed to the study design. CDS performed the data analysis and interpretation under the supervision of CvB. CDS and MSL drafted the manuscript, and KO and CvB provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgement

This work was supported by a grant from the Swiss National Science Foundation (No. 100014_146074) to Klaus Oberauer.

Abstract

Self-control is an important factor for a wide range of positive life events. The strength-model assumes that self-control depends on a limited domain-general resource. Acts of self-control exert this resource, resulting in a state of impaired self-control known as ego-depletion. Although a large body of literature appears to confirm the predictions of the strength-model, recent failed replication attempts and meta-analyses suggest that ego-depletion effects are not as robust as assumed. By reexamining four previously reported data sets derived from cognitive test batteries, we tested two predictions derived from the strength-model: 1) ego-depletion yields a performance decline over multiple tasks, 2) ego-depletion effects are found for all cognitive abilities requiring self-control. Most of the Bayesian hypothesis tests provided evidence in favor of null effects, resulting in the rejection of the hypotheses. Thus, the current reanalyses show that the strength model does not contribute to our understanding of performance in cognitive tasks.

Keywords

Self-control, ego-depletion, working memory

3.1 Introduction

Self-control refers to the ability to control thoughts, emotions, and impulses in order to stay focused on aversive and cognitively demanding tasks (Inzlicht, Schmeichel, & Macrae, 2014). Self-control failure has been connected to various health-related issues such as physical inactivity, obesity, and substance abuse (Baumeister & Heatherton, 1996). To explain these effects, Baumeister and colleagues postulated the strength-model of self-control (Baumeister, 2014; Baumeister & Heatherton, 1996; Baumeister, Vohs, & Tice, 2007). This model is based on the assumption that self-control depends on a limited, domain-general resource that becomes exerted by use. The exertion leads to a temporary state of "ego-depletion" during which further acts of self-control are more likely to fail. These performance impairments are assumed to become more severe and more difficult to counteract the less of the resource is available (Baumeister, 2014; Vohs, Baumeister, & Schmeichel, 2013). Furthermore, as the resource is thought to be shared among processes, performance impairments are predicted to affect a wide range of behavior such as emotion regulation, planning, and working memory (Schmeichel, 2007; Webb & Sheeran, 2003).

Ego-depletion effects are typically studied with a sequential two-task paradigm. In an initial manipulation task, participants engage in an effortful, self-control exerting condition, or a control condition not requiring self-control. The strength-model predicts worse performance in the subsequent outcome task for participants in the self-control exerting condition compared to those in the control condition. Indeed, more than 200 reported experiments found such ego-depletion effects (cf. Hagger, Wood, Stiff, & Chatzisarantis, 2010).

However, the initial meta-analytic estimate of a moderate ego-depletion effect (*d* = 0.62; Hagger, Wood, Stiff, & Chatzisarantis, 2010), has recently been questioned after a series of meta-analytic tests (Carter, Kofler, Forster, & McCullough, 2015; Carter & McCullough, 2014) and failed replication attempts (e.g., Carter & McCullough, 2013b; Murtagh & Todd, 2004; Xu et al., 2014) including a preregistered study (Lurquin et al., 2016) and a Registered Replication Report involving 23 labs (Hagger et al., 2016). Further unpublished null-findings may exist.

Furthermore, two aspects of the strength-model have so far received little attention: the extent of ego-depletion over time, and the domain-generality of ego-depletion effects. First, given that self-control is assumed to depend on a limited resource pool, its exertion across several successive tasks should lead to extensive ego-depletion along with a severe decline in performance. However, the two-task paradigm has only rarely been extended to incorporate multiple manipulation tasks. One exception is a study by Vohs, Baumeister, and Schmeichel (2013), in which the authors reported increasingly severe performance impairments after three or four tasks (but see Carter et al., 2015).

Second, in line with the assumption of domain-generality, exertion of the common resource should affect performance in any outcome task involving self-control. Ego-depletion effects have indeed been reported for a wide range of cognitive tasks. For example, after an attention control task, participants performed worse in logic and reasoning (Schmeichel, Vohs, & Baumeister, 2003) or in an operation span task (Schmeichel, 2007). However, the two-task paradigm with one outcome task allows for testing single indicators and abilities only. Hence, ego-depletion has not been tested yet across different abilities within the same set of participants or across multiple tasks measuring the same ability.

Current Study

We reanalyzed four studies where participants were engaged in test batteries of cognitively demanding tasks for 2.5 to 4.5 hours (Herkert, 2012; von Bastian & Eschen, 2016; von Bastian & Oberauer, 2013; von Bastian, Souza, & Gade, 2016). Reexamining these data has two advantages. First, it allows for estimating ego-depletion effects across a series of multiple cognitive tasks. Second, the cognitive tasks administered in these studies are more varied than in previous ego-depletion studies, thereby providing an opportunity to test whether the phenomenon extends to any effortful task or only to a subset of tasks. For example, although the strength-model suggests that ego-depletion affects executive control in general, tasks measuring the ability to shift mental task-sets have not yet been included as depleting or outcome tasks. The current study fills these gaps by testing for ego-depletion effects across multiple tasks measuring multiple cognitive abilities, and administered in four studies with moderate to relatively large sample sizes. Based on the strength-model of self-control, our hypotheses are as follows:

1. *Performance decline:* Performance in tasks demanding self-control is worse at the end than at the beginning of a series of tasks.
2. *Generality of the ego-depletion effect:* Prior self-control exertion negatively affects subsequent performance on any other task requiring self-control independent of the specifics of the task and the ability measured by it.

3.2 Methods

In the present study, data sets of four previous studies from members of our laboratory were reexamined. We included all data sets that were available at the time of planning the current project (i.e., October 2013) and that entailed test batteries consisting of multiple cognitive tasks administered in two versions with reversed task orders but otherwise identical stimuli. Three of the studies investigated the effectiveness of computer-based working memory training (Herkert, 2012; von Bastian & Eschen, 2016; von Bastian & Oberauer, 2013) and one evaluated possible bilingual advantages across multiple cognitive functions (von Bastian et al., 2016). One training study was a master thesis (Herkert, 2012) and the other three studies are published in peer-reviewed journals.

3.2.1 Participants

Sample sizes in the four reexamined studies ranged from 57 to 129 mentally and physically healthy young adults (see Table 1 for detailed participant descriptions). All participants were recruited from the participant pools of the Department of Psychology of the University of Zurich and through advertisements posted at the University's campus. Participants gave written informed consent and received either financial compensation or course credits.

3.2.2 Studies and Materials

Three training studies investigated the effectiveness of computer-based working memory training. Participants completed 20 intensive training sessions over the course of four weeks. Cognitive abilities (e.g., reasoning and executive functions) were assessed on a broad set of tasks before and after training to measure transfer effects to related cognitive functions. For the current study, analysis was restricted to pre-assessment data, because participants might have been motivated to show training effects in post-assessment, which could have counteracted effects of self-control exertion (e.g., Inzlicht et al., 2014). Pre-assessment data was collected in one lab session that lasted between 2.5 (Herkert, 2012) and

4.5 hours (von Bastian & Oberauer, 2013), depending on the number of tasks. The fourth study focused on individual differences and evaluated a possible bilingual advantage in cognitive functions. Participants completed various tasks measuring nine cognitive abilities in a single 4.5-hr session (von Bastian et al., 2016).

A maximum of five participants (four in the study by von Bastian & Eschen, 2016) were tested simultaneously. The test batteries were divided into several blocks (see Table 1 for an overview of the characteristics of test batteries). Between blocks, participants had a break of 5 min (Herkert, 2012; von Bastian & Eschen, 2016; von Bastian et al., 2016) or 10 min (von Bastian & Oberauer, 2013) during which they could eat and drink.

Half of the participants completed the test battery in reversed order (i.e., they started with the forward order's last task of the last block and ended the test session with the first task of the first block). A description of the paradigms used in the test batteries is provided in Table 2.

Table 1

Sample Demographics and Test Battery Characteristics for Each Study

Study Feature	Study			
	von Bastian & Oberauer (2013)	Herkert (2012)	von Bastian & Eschen (2016)	von Bastian, et al. (2016)
Sample				
Sample Size (*N*)	121	57[a]	129[a]	118
Gender (f/m)	83/38	30/26	93/37	74/44
Age ($M \pm SD$)	23.34 ± 3.92	23.00 ± 3.30	22.92 ± 3.30	24.17 ± 3.62
Test Battery				
Duration (h)	4.5	2.5	3.5	4.5
No. of Blocks	5	4	3	4
No. of Tasks	16	10	12	20

[a] The original data set contained one additional participant that was excluded from the current analyses due to missing data in one task.

Table 2

Paradigms Administered in the Included Studies

Paradigm	Description	DV
	Working Memory	
Brown-Peterson[a, c, d]	Memorize a list of items, followed by a distractor task. At the end of each trial, recall previously memorized items in serial order.	a
Complex span[b, c, d]	Alternately, memorize an item and solve a distractor task. At the end of each trial, recall previously memorized items in serial order.	a
Keep-Track[d]	Memorize a set of items. Subsequently update individual memoranda when new information is displayed.	a
Local recognition[a, b, c]	Memorize the locations of sequentially presented words. In the subsequent recognition phase, decide whether the locations of the probe words match those during memorization.	a
Memory updating[a, c]	Memorize a set of digits. Subsequently update individual memoranda by simple arithmetic operations.	a
Monitoring[a, d]	Identify relations between objects in a matrix of changing objects.	a
N-back[b]	For a set of sequentially presented items, decide whether the current item matches the one *n* steps back.	a
Relational integration[a,b]	Infer the resulting relationship of individually presented items.	a
Spatial short-term memory[d]	Memorize and later reproduce spatial locations of dots in a 10-by-10 grid.	a
	Reasoning	
Berlin intelligence-structure test - short version (BIS-4S)[a]	Intelligence test measuring reasoning, creativity, memory, and speed in the verbal, numerical, and spatial-figural domain.	a
Bochum matrices test advanced - short version (BOMAT)[b]	Complete matrix patterns by choosing one of six response options.	a
Diagramming relationships[c, d]	Select one out of five diagrams that best describes the relationship between three nouns.	a

Letter sets[c, d]	Discover the logical pattern underlying sets of four letters to identify the deviating letter set.	a
Locations test[c, d]	In four rows of dashes, one dash is replaced by an "x" according to a particular rule. Infer the position of the "x" in the fifth row.	a
Syllogisms[a, c, d]	Determine the valid conclusion from two related premises.	a
Raven's advanced progressive matrices (APM)[c, d]	Complete matrix patterns by choosing one of eight response options.	a
	Shifting	
Task switching[a, b, d]	Classify bivalent stimuli according to one of two categorization rules indicated by a cue.	s
	Inhibition	
Letter-flanker[a, d]	Categorize a centrally presented letter that is flanked by three congruent or incongruent letters as vowel or consonant.	s
Number-Stroop[a, d]	Determine the number of digits instead of their value.	s
Simon[d]	React to the color of a circle presented in the left or right half of the screen instead of its spatial location.	s
	Other	
Quiz[b, c]	Answer general knowledge questions.	a
Face matching[a]	Decide whether two simultaneously presented faces show the same or different persons.	s

Note. Detailed descriptions of the tasks can be found in the Supplemental Material available online (Table S1 and S2) and in the original publications (von Bastian & Eschen, 2016; von Bastian & Oberauer, 2013; von Bastian et al., 2016). DV = dependent variable; a = accuracy-based dependent variable; s = speed-based dependent variable.

[a] von Bastian and Oberauer (2013)
[b] Herkert (2012)
[c] von Bastian and Eschen (2016)
[d] von Bastian et al. (2016)

3.3 Data Analysis

For the present study, all dependent variables were *z*-transformed to allow for comparisons across accuracy- and speed-based outcome measures (see Table 2). *Z*-transformation was done across both task orders (forward and backward). Therefore, effects of task order on task performance are reflected in a difference in *z*-scores. For example, if participants in the forward condition outperformed those in the backward condition on a given task, the averaged *z*-score for that task will be positive for the forward condition and negative for the backward condition.

Data were analyzed with Bayes statistics. In contrast to frequentist statistics, Bayes statistics have the advantage that they allow for quantifying evidence for both the alternative (H_1) hypothesis and the null (H_0) hypothesis. Strength of evidence is expressed by the Bayes factor (*BF*), which is the probability of the data under one hypothesis divided by the probability of the data under the other hypothesis (Jeffreys, 1961). A *BF* of 1 corresponds to perfect ambiguity, which means that both hypotheses are equally supported, a *BF* smaller than 1 reflects evidence for the hypothesis in the denominator (typically H_0), and a *BF* greater than 1 evidence in favor of the hypothesis in the numerator (typically H_1). For example, a *BF* of 5 means that the data are five times more likely under the hypothesis in the numerator compared to the hypothesis in the denominator.

All analyses were conducted in R (R Core Team, 2015) with the “BayesFactor” package (Morey & Rouder, 2015). Bayes Factors were computed using the default prior settings (i.e., Cauchy distribution with a medium scaling factor) of the “BayesFactor” package and interpreted according to the guidelines (see Table 3) suggested by Wetzels and Wagenmakers (2012). Figure 2 shows how these verbal labels translate into probabilities.

However, these guidelines must not be understood as thresholds for inferential decisions (cf. Morey, 2015); rather they provide verbal labels for ranges of Bayes Factors to facilitate their interpretation. For convenience, *BF*s below 1 (i.e., *BF*s in favor of the H_0) are expressed as 1/*BF*. The data and the R-scripts are available at the Open Science Framework (https://osf.io/5s84n/).

Table 3

Verbal Labels for Interpreting Bayes Factors

Bayes Factor	Label
1 to 3	Anecdotal
3 to 10	Substantial
10 to 30	Strong
30 to 100	Very strong
> 100	Decisive

Note. Adapted from Wetzels and Wagenmakers (2012).

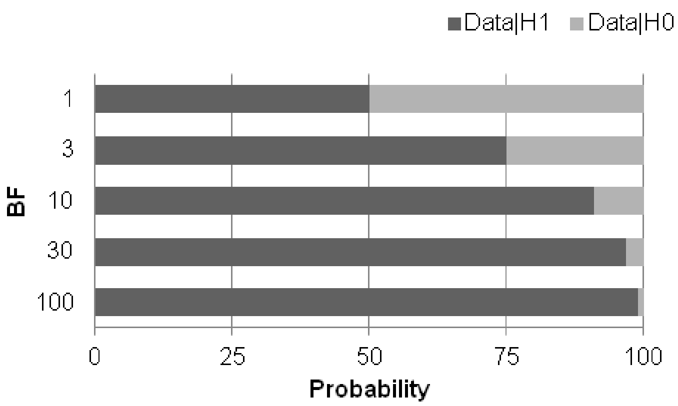

Figure 1. Probability of the data under the alternative hypothesis (H1) relative to the null hypothesis (H0) in correspondence to the labels suggested by Wetzels and Wagenmakers (2012). For example, a *BF* of 3 favoring H1 corresponds to the data being 75 % more likely under H1 than under H0.

Task-focused comparisons. To test Hypothesis 1: *performance decline*, we carried out planned comparisons between the two task-order groups for each task. We conducted one-tailed two-sample Bayesian t-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) using the ttestBF function to compare performance on each task in forward and backward order (see Table 4 for an overview). Based on the strength-model's predictions, participants in the forward condition should perform better in the first half of the test battery than those in the backward condition who completed the same tasks as the second half of the test battery. Conversely, the opposite pattern should emerge for the tasks participants in the forward condition performed in the second half of the test battery.

As the strength model postulates that self-control resources regenerate after periods of rest (cf. Baumeister & Heatherton, 1996), we focused on potential ego-depletion effects within each block, consisting of 45 to 60 minutes of cognitive tasks without a break, and compared performance on the first versus the last task in each block (cf. Hypothesis 1: *performance decline*).

To evaluate whether ego-depletion effects could be observed for any cognitive ability requiring self-control (Hypothesis 2: *task-independent effects*), we examined change in performance in relation to whether tasks were completed early compared to late in the test battery. These analyses were carried out separately for tasks measuring inhibition, working memory, reasoning, and shifting. *BF*s in favor of H_0 indicate the absence of ego-depletion effects or effects in the opposite direction from ego-depletion; *BF*s in favor of H_1 indicate evidence for the presence of ego-depletion effects.

Small-scale meta-analyses. To evaluate the evidence gathered from the analyses outlined above across the four studies, we additionally conducted small-scale meta-analyses (cf. Cumming, 2014) using the meta.ttestBF function with a one-sided interval. To avoid possible biases from non-independence of effects (cf. Rosenthal & Rubin, 1986), we averaged the *t*-values gathered from the task-focused comparisons within each study.

3.4 Results

3.4.1 Hypothesis 1: Performance Decline

First, we focused on the pairwise comparisons of the tasks completed first in each block in one task order, and last in their block in the other task order, because we expected that ego-depletion effects should be strongest for these tasks: Performance should be worse when the task was attempted at the end of a block than at the beginning. Out of 22 comparisons (in the studies of von Bastian & Oberauer, 2013, and Herkert, 2012, the first block comprised only one task and was therefore excluded from the analysis) three showed anecdotal evidence in favour of an ego-depletion effect (i.e., the face matching task in the study by von Bastian & Oberauer, 2013, the letter integration task in the study by Herkert, 2012, and the numerical complex span task in the study by von Bastian & Eschen, 2016, cf. Table 4). However, the meta-analysis across blocks and studies revealed a *BF* = 3.03 in favour of the H_0 (cf. Figure 2). Overall, there is substantial evidence for the absence of self-control exertion within blocks.

3.4.2 Hypothesis 2: Generality of the Ego-Depletion Effect

To test the second hypothesis that effects of self-control exertion emerge regardless of the cognitive ability measured by the task, we investigated the effects separately for tasks measuring inhibition, working memory, reasoning, and shifting. Across all studies, inhibition was measured five times by three tasks (i.e., letter-flanker, number Stroop, and Simon task). Bayesian t-tests revealed BFs between 3.73 and 13.16, and a meta-analysis across all inhibition tasks a BF = 12.67 in favor of the H0 (cf. Table 4 and Figure 2). Again, these results do not support the ego-depletion hypothesis. For example, the Stroop task, which is one of the most frequently used outcome tasks, showed no evidence in favor of the strength-model (BF = 7.30 and BF = 13.16 for H0 in the studies of von Bastian & Oberauer, 2013 and von Bastian et al., 2016, respectively).

Examining the overall twenty-nine working memory tasks revealed a similar pattern. The meta-analysis across all WM tasks yielded a strong BF = 23.84 in favor of the H0 contradictory to the strength-model's prediction (cf. Figure 2). In ego-depletion research, complex span paradigms are typically used to assess working memory. However, even with a scope restricted to the six comparisons for complex span tasks, we found anecdotal evidence for an ego-depletion effect only once (BF = 1.21; numerical complex span task in von Bastian & Eschen, 2016), whereas the other five comparisons returned evidence for the H0 (BF from 2.78 to 8.77). A meta-analysis across the complex span tasks yielded a BF = 8.97 in favor of the H0. To conclude, the balance of evidence speaks against an ego-depletion effect in working memory tasks.

Finally, we investigated whether two domains less frequently examined for ego-depletion effects, reasoning and task shifting, are prone to ego-depletion. Reasoning was measured with 13 tasks. Results showed eight pairwise comparisons with a BF > 3. Evidence favored H1 for one task only (i.e., Raven's APM in the study of von Bastian & Eschen, 2016) with a BF = 8.66, indicating that participants performed better in this task when doing it in the first half than in the second half of the test battery. Raven's APM was also used in the study of von Bastian et al. (2016), where the pairwise comparison however yielded substantial evidence for the H0 (BF = 6.17). A meta-analysis across all reasoning tasks revealed substantial evidence for the absence of an ego-depletion effect, with a BF = 5.19. Shifting was measured with eight tasks across the four studies. Pairwise comparisons for these tasks yielded substantial BFs ranging between 5.88 and 14.29 and a strong meta-analytical BF = 22.31 in favor of H0 (cf. Table 4 and Figure 2).

For the sake of completeness, we also examined the three control tasks (labeled "other tasks" in Figure 2) not measuring any of the above mentioned cognitive functions. Only one of them showed a BF > 3, namely the Quiz in the study of von Bastian and Eschen (2016) with a BF = 3.62 in favor of the H0. There was again no evidence for an ego-depletion effect.

Taken together, we found convincing evidence for the absence of ego-depletion effects for each of the cognitive abilities tested. Neither did tasks typically used in ego-depletion experiments (i.e., Stroop task and complex span paradigm) nor did tasks assessing related cognitive functions (i.e., shifting and reasoning) indicate performance impairments after completing several cognitively demanding tasks.

Table 4

Comparisons of Z-Transformed Means and Standard Deviations between Administration Orders for Each Task

		Forward		Backward		Evidence	
Task	Material	*M*	*SD*	*M*	*SD*	H	*BF*
von Bastian & Oberauer (2013)							
Block 1							
BIS-4S	M	-0.08	1.02	0.08	0.98	H_0	9.09
Block 2							
Monitoring	V	-0.33	0.96	0.33	0.94	H_0	23.81
Brown-Peterson	F	-0.36	0.98	0.36	0.89	H_0	25.64
Letter-flanker	V	-0.02	0.93	0.01	1.08	H_0	5.81
Memory updating	N	-0.16	0.98	0.16	1.01	H_0	13.33
Block 3							
Local-recognition	V	-0.07	0.94	0.07	1.06	H_0	8.26
Task switching	F	-0.16	1.03	0.16	0.95	H_0	13.51
Brown-Peterson	V	-0.19	1.02	0.18	0.95	H_0	14.93
Face matching	**F**	**-0.15**	**0.97**	**0.15**	**1.02**	$\mathbf{H_1}$	**1.31**
Block 4							
Task switching	N	0.10	1.00	-0.10	1.00	H_0	10.10
Syllogisms	V	-0.07	0.98	0.07	1.03	H_0	2.58
Monitoring	N	-0.09	1.00	0.09	1.00	H_0	1.93
Block 5							
Brown-Peterson	N	0.29	0.99	-0.28	0.93	H_0	20.83
Task switching	V	0.18	0.98	-0.17	1.00	H_0	14.29
Monitoring	F	-0.03	1.02	0.03	0.99	H_0	4.13
Number Stroop	N	0.05	1.02	-0.05	0.99	H_0	7.30
Herkert (2012)							
Block 1							
BOMAT	F	0.10	0.93	-0.08	1.09	H_0	2.10
Block 2							
Letter integration	**V**	**0.16**	**0.86**	**-0.21**	**1.13**	$\mathbf{H_1}$	**1.09**
Task switching	V	-0.09	1.04	0.10	0.97	H_0	5.88
Complex span	N	-0.21	1.02	0.22	0.96	H_0	8.77
Block 3							
Kinship integration	V	-0.04	0.95	0.00	1.06	H_0	4.13
Quiz	V	-0.08	0.98	0.06	1.04	H_0	2.46
nBack	V	0.07	1.02	-0.04	1.00	H_0	4.85
Block 4							
Local-recognition	V	0.23	1.05	-0.30	0.87	H_0	10.20
Task switching	F	0.11	1.03	-0.13	0.97	H_0	6.41

Complex span	V	-0.07	1.09	0.03	0.90	H_0	2.78
von Bastian & Eschen (2016)							
Block 1							
Complex span	**N**	**0.15**	**1.03**	**-0.14**	**0.95**	**H_1**	**1.21**
Diagramming relationships	V	-0.05	0.95	0.06	1.06	H_0	8.00
Local-recognition	V	-0.07	1.03	0.08	0.98	H_0	9.01
Brown-Peterson	V	-0.12	1.09	0.12	0.89	H_0	11.63
Block 2							
Syllogisms	V	0.00	0.96	0.00	1.06	H_0	5.41
Memory updating	N	0.06	0.95	-0.06	1.06	H_0	2.89
Quiz	V	-0.03	0.95	0.05	1.05	H_0	3.62
Complex span	F	0.06	0.98	-0.06	1.03	H_0	8.40
Block 3							
Locations test	F	0.02	0.96	-0.02	1.05	H_0	6.21
Complex span	V	-0.04	1.09	0.04	0.91	H_0	3.73
Ravens APM	**F**	**-0.20**	**1.03**	**0.25**	**0.88**	**H_1**	**8.66**
Letter sets	V	0.01	0.93	-0.01	1.09	H_0	5.81
von Bastian et al. (2016)							
Block 1							
Ravens APM	F	-0.02	0.99	0.02	1.01	H_0	6.17
Task switching	N	-0.11	0.97	0.11	1.02	H_0	10.31
Simon	F	0.04	0.97	-0.04	1.03	H_0	3.73
Monitoring	N	-0.09	0.94	0.09	1.06	H_0	9.26
Keep-track	V	-0.30	0.97	0.30	0.95	H_0	21.28
Block 2							
Brown-Peterson	V	-0.09	1.03	0.09	0.97	H_0	9.52
Letter-flanker	V	-0.14	1.16	0.14	0.80	H_0	11.76
Monitoring	F	-0.16	1.02	0.16	0.97	H_0	12.99
Keep-track	N	-0.02	0.95	0.02	1.05	H_0	6.10
Task switching	F	-0.10	1.07	0.10	0.92	H_0	9.80
Syllogisms	V	-0.09	1.02	0.09	0.98	H_0	2.11
Block 3							
Keep-track	F	0.07	0.94	-0.07	1.06	H_0	8.47
Task switching	V	0.04	1.09	-0.04	0.91	H_0	6.67
Letter sets	**V**	**-0.17**	**1.18**	**0.17**	**0.75**	**H_1**	**1.72**
Number Stroop	N	0.16	0.94	-0.16	1.04	H_0	13.16
Spatial short-term memory	F	0.05	0.93	-0.05	1.08	H_0	7.30
Block 4							
Locations test	F	0.02	0.91	-0.02	1.09	H_0	6.13
Complex span	N	0.03	1.01	-0.03	1.00	H_0	6.37
Monitoring	V	0.01	1.12	-0.01	0.87	H_0	5.46
Diagramming relationships	V	-0.13	0.98	0.13	1.01	H_0	1.20

Note. Tasks are listed in forward order. Tasks which provided evidence in favor of ego-depletion are printed bold. Sample sizes for forward order were *n* = 60, 30, 66, and 59, and for backward order *n* = 61, 27, 63, and 59 for the studies of von Bastian and Oberauer (2013), Herkert (2012), von Bastian and Eschen (2016), and von Bastian et al. (2016), respectively. H = hypothesis supported by data; H_0 = null hypothesis (i.e., no performance differences or effect in opposite direction from ego-depletion); H_1 = alternative hypothesis (i.e., ego-depletion effect; *BF* = Bayes Factor in favor of supported hypothesis (H); F= figural; M = mixed; N = numerical; V = verbal stimulus material; APM = Advanced Progressive Matrices; BIS-4S = Berlin Intelligence-Structure Test; BOMAT = Bochum Matrices Test).

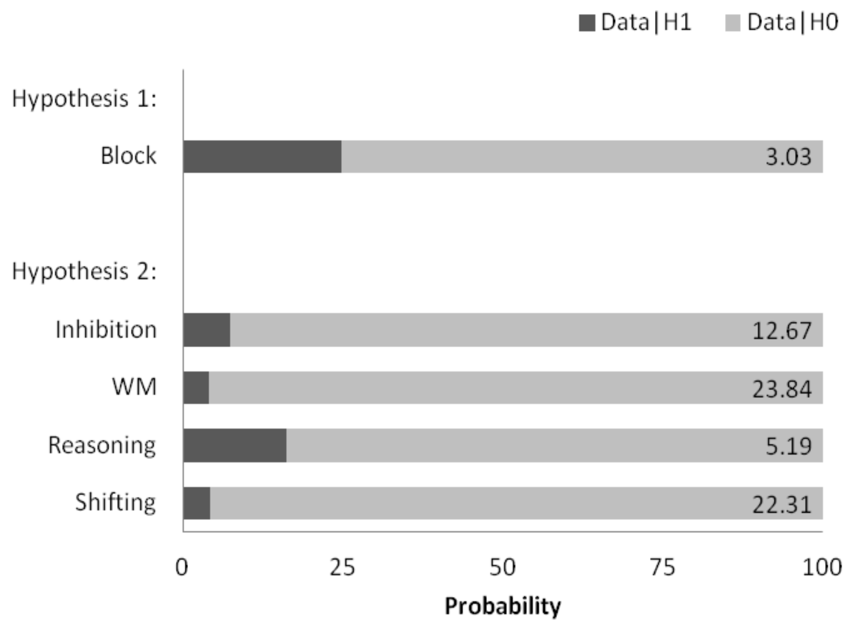

Figure 2. Probability of data under H_1 relative to H_0 for the meta-analytical *BF*s for Hypothesis 1 (*performance decline*) and Hypothesis 2 *(generality of the ego-depletion effect*). Bars are labeled with the meta-analytic *BF*s. H1 = alternative hypothesis (i.e., ego-depletion effect), H0 = null hypothesis (i.e., no performance differences, or effect in the opposite direction of ego depletion).

3.5 Discussion

We analyzed four previously reported data sets with larger-than-typical sample sizes and multiple cognitive tasks to test two predictions of the strength-model, (1) performance is worse after working on a series of other tasks than at the beginning of the series and (2) ego-depletion is observed for all cognitive abilities that require self-control. Based on Bayesian inference, both hypotheses were disconfirmed: With few exceptions, all comparisons of performance on the same task when attempted first in a block and when attempted last, after 45 minutes or more of other demanding cognitive tasks, provided evidence against an ego-depletion effect.

Next, we examined ego-depletion effects on inhibition, working memory, reasoning, and shifting tasks separately. The evidence was unambiguous, with meta-analytic *BF*s between 5.19 and 23.84, indicating that the data were at least five times more likely under H_0 than H_1. Restriction to tasks typically used in ego-depletion research (Stroop and complex span tasks) also revealed little evidence for ego-depletion and instead supported the H_0. Taken together, our findings contradict the notion of a limited and domain-general self-control resource as assumed by the strength-model.

3.5.1 Strengths

Our study has five major strengths: the relatively large sample sizes, the reanalysis of existing data sets, the examination of a wide range of cognitive tasks, the use of the same tasks for manipulation and measurement of ego-depletion, and the Bayesian evaluation of evidence.

First, in contrast to previous ego-depletion studies with an average group size of n = 27 (Carter et al., 2015) yielding relatively weak statistical evidence, 3 out of the 4 studies we reanalyzed featured $n \geq 59$ participants per group. These sample sizes provided sufficiently strong data to yield substantial, and in most cases strong, evidence to adjudicate between the competing hypotheses, as reflected in the size of the Bayes Factors. The relatively large sample sizes render it also unlikely that our findings are based on a disproportionally high ratio of participants who are not or less susceptible to ego-depletion effects (cf. Job, Dweck, & Walton, 2010).

Second, the studies we reexamined were originally not intended to test ego-depletion, thus providing a good opportunity to test the effects in the absence of experimenter effects potentially biasing effects in one or another direction. As Rosenthal and colleagues showed in multiple experiments, the expectancies of an experimenter can influence participants through verbal and nonverbal behavior to the extent that they act as expected (Rosenthal, 2002). Baumeister (2016) alluded to this possibility when he wrote that "Most of my advisees have been able to produce such effects [i.e., ego-depletion effects] (...) A few of them have not been able to replicate the basic effect after several trials (...) some people simply seem to lack whatever skills and talents are needed" (p. 156). The experimenters carrying out our studies were skilled in administering cognitive tasks, but were not trained to produce ego-depletion effects.

Third, analyzing a sequence of more than two tasks circumvents one of the major issues in ego-depletion experiments. As the length of the initial manipulation task has been found to moderate the ego-depletion effect (Hagger et al., 2010), insufficiently strong demands of that initial task have been discussed as a cause of failures to detect ego-depletion. Completing cognitively demanding tasks for more than 45 min should, however, be sufficient to induce strong and reliable ego-depletion. Moreover, the extension of the classic two-task paradigm allows for examining not only single indicators but a set of cognitive tasks and abilities.

Finally, basing our conclusions on Bayesian inference has the major advantage that it not only allows for quantifying evidence for H1, but also in favor of H0. Following the guidelines of Wetzels and Wagenmakers (2012; cf. Table 3), Bayesian meta-analyses across studies showed at least substantial evidence in favor of H0. In sum, it is unlikely that the absence of depletion effects in our study is due to insufficient data, random fluctuations in depletion susceptibility, experimenter effects, or task selection.

3.5.2 Limitations

Several limitations of the current study have to be considered. First, no manipulation check data was collected. Therefore, we do not know how engaging in the test batteries affected participants' subjective sense of motivation, emotions, or fatigue over time. Baumeister and colleagues (2007) described several processes (e.g., positive emotions, cash incentives, and motivation) possibly counteracting ego-depletion. For example, Muraven and Slessareva (2003) showed that highly motivated participants put more effort in processing the outcome task and therefore could compensate for loss of self-control resources (for similar results see Luethi et al., 2016). However, given the high cognitive demands and long duration of the test batteries, it seems unlikely that processes such as consistently increasing motivation and positive emotions resulted in a null effect.

Second, participants could take breaks (5-10 min) in between blocks (45-60 min). The strength model postulates that self-control resources regenerate after periods of rest (cf. Baumeister & Heatherton, 1996), but it is yet unclear whether such recovery effects exist (Hagger et al., 2010), and how long recovery after such extensive periods of depletion would take. If short breaks enable recovery, they could have mitigated ego-depletion effects across blocks (as used for our tests of hypothesis 2) but not within blocks (as used for testing hypothesis 1).

Third, the current work focused almost entirely on computer-based cognitive tasks (except the paper-and-pencil BIS-4S test). Baumeister and Vohs (2016) recently criticized the adequacy of computerized tasks to test ego-depletion. Moreover, they argued that cognitive tasks in general are unsuitable to detect ego-depletion effects as they are only indirectly affected by self-control. However, previous studies found ego-depletion in both cognitive (e.g., Schmeichel, 2007) and computerized tasks (Sripada, Kessler, & Jonides, 2014). From a theoretical perspective, it remains unclear why ego-depletion effects should not be expected for computer-based cognitive tasks.

3.5.3 Implications and Future Directions

Taken together, ego-depletion seems to be a phenomenon that, if existing, is less pervasive than expected, and depends on several boundary conditions. The strength-model does not make assumptions about such boundary conditions. The null-findings cast substantial doubt on the strength-model of self-control in its current form.

An alternative interpretation of our findings is that learning processes counteracted ego-depletion. Earlier work on learned industriousness (e.g., Eisenberger, 1992) and adaptation-level theory (Helson, 1964) suggests adaptation effects after self-control effort, resulting in greater subsequent self-control exertion and, thus, better performance. Indeed, some recent studies reported such enhanced self-regulation after self-control exertion (Carter & McCullough, 2013a; Converse & Deshon, 2009; Dewitte, Bruyneel, & Geyskens, 2009; Xu et al., 2014). However, a large body of evidence from cognitive training studies showing only weak transfer effects even after weeks of intensive task practice (for a meta-analysis, see Melby-Lervåg, Redick, & Hulme, 2016) strongly contradicts the notion that self-control can be improved to a meaningful degree by merely completing a test battery. Still, future research is required that systematically manipulates the duration of self-control exertion to disentangle the two theoretical accounts.

3.6 Conclusion

After years of ego-depletion research, there is a growing number of recent studies questioning the robustness of the effect (Hagger et al., 2016; Lurquin et al., 2016; Xu et al., 2014). This reanalysis was conducted in response to calls for further large-scale tests of the ego-depletion effect with task conditions ensuring self-control exertion. Yet, we found hardly any evidence in favor of ego-depletion. Although the size of the ego-depletion effect and its boundary conditions are still to be explored, the current work adds to the notion that the magnitude and the robustness of ego-depletion have been overestimated in previous research.

3.7 Supplementary Material

Table S1

Paradigms Implemented in the Study by Herkert (2012)

Paradigm	Description
	Working memory
Complex span	This paradigm consists of a memory and a distractor task which are alternately presented. First, a memory item is displayed for 1000 ms, followed by a distractor task that has to be solved within 3000 ms. Immediately after a response, the next memorandum is displayed. After 3 to 7 memory-distractor combinations the memoranda had to be recalled in serial order. Performance was measured as the percentage of correctly recalled memory items. In the **numerical task version**, participants had to remember two-digit numbers and evaluate the correctness of simple arithmetic equations. In the **verbal task version**, participants memorized words instead of numbers and judged the semantic plausibility of sentences.
Local recognition	In each trial, the positions of 2 to 5 sequentially presented words had to be memorized. During the recognition phase, participants had to decide whether probe words matched the word at this position. Probes could be of three types, positive probes, in which the word is presented at the correct location, negative probes in which a new word is presented, and intrusion probes in which a word is displayed at the wrong position. The discrimination parameter *d*' (from signal detection theory) served as score. It is calculated from the z-transformed relative frequency of hits minus the z-transformed relative frequency of false alarms to intrusion probes.
n-back	A list of letters was sequentially presented on the screen. Participants had to decide whether the current letter has already been shown *n* steps ago. The letter list was intermixed with intrusions, that is, target letters displayed at the wrong position (*n*-1 or *n*+1). *N* varied between 2 and 4. As in the local recognition task, the discrimination parameter *d*' was used as dependent variable (Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011).
Relational integration	In this paradigm, sequentially presented items have to be mentally integrated into a superordinate structure. The number of correct answers was used as dependent variable (von Bastian & Oberauer, 2013). In the **letter integration task**, letters were sequentially presented in a row of place holders in random order. After the last letter disappeared participants had to indicate whether the presented letters formed a word. In the **kinship integration task**, 2 to 4 short sentences describing a

	kinship relation between two persons (e.g., "Jeff is Sarah's brother"; "Michael is Jeff's son") were presented. Afterwards participants were asked to decide how two of the before mentioned persons are related to each other (e.g., "Michael is Sarah's?" - "Nephew" would be the correct answer).
	Reasoning
BOMAT	In this matrix reasoning test, participants had to logically complete a pattern in a 3-by-5-matrix selecting 1 out of 6 response options. The proportion of correctly solved items served as dependent variable.
	Shifting
Task switching	The task switching paradigm comprised the classification of bivalent stimuli according to two categorization rules which switch after every second trial. The ability to shift between two classification rules was operationalized with reaction time (RT) switch costs which were calculated by subtracting RTs in task switch trials from RTs in task repetition trials. In the **verbal task version**, participants had to switch between deciding whether a word denoted a city or a river, or whether it was written in blue or orange. Colored geometric shapes were used as stimuli for the **figural task version**. The categorization rules were to classify the shape as round or angular, or as blue or green.
	Other
Quiz	A set of 16 general knowledge questions served as control task. The questions came from the field of movie, literature, cartoon, contemporary music, and language. Participants had to type their answers and confirm them by pressing enter.

Table S2

References for a Detailed Description of Implemented Paradigms in the Included Studies

Paradigm	References
Brown-Peterson[a, c, d]	Brown (1958)
Complex span[b, c, d]	Engle and Turner (1987)
Keep-track[d]	Miyake et al. (2000)
Local-recognition[a, b, c]	Oberauer (2005)
Memory-updating[a, c]	Oberauer (2006)
Monitoring[a, d]	Oberauer, Süß, Wilhelm, and Wittman (2003)
N-back[b]	Cohen et al. (1997)
Relational integration[a,b]	von Bastian and Oberauer (2013)
Spatial short-term memory[d]	Lewandowsky, Oberauer, Yang, and Ecker (2010)
Berlin intelligence-structure test - short version (BIS-4S)[a]	Jäger, Süss, and Beauducel (1997)

Bochum matrices test advanced - short version (BOMAT)[b]	Hossiep, Hasella, and Turck (2001)
Diagramming relationships[c, d]	Ekstrom, French, Harman, and Dermen (1976)
Letter sets[c, d]	Ekstrom et al. (1976)
Locations test[c, d]	Ekstrom et al. (1976)
Syllogisms[a, c, d]	Ekstrom et al. (1976)
Raven's advanced progressive matrices (APM)[c, d]	Raven (1990)
Task switching[a, b, d]	Monsell (2003)
Letter-flanker[a, d]	Eriksen and Eriksen (1974)
Number-Stroop[a, d]	Salthouse and Meinz (1995)
Simon[d]	Simon (1990)
Quiz[b, c]	von Bastian and Eschen (2016)
Face matching[a]	von Bastian and Oberauer (2013)

Note
[a] von Bastian and Oberauer (2013).
[b] Herkert (2012).
[c] von Bastian and Eschen (2016).
[d] von Bastian, Souza, and Gade (2016).

4 General Discussion

4.1 Summary

In the empirical part of this thesis a large-scale training study and a reanalysis of four existing data sets is reported. In the first study (No Evidence for Effects of Updating and Binding Training on Working Memory Capacity and Efficiency), we intended to disentangle improvements in WM capacity and efficiency after updating and binding training. Therefore, we evaluated intervention effects on structurally different WM tasks (near transfer) and closely related abilities such as reasoning, processing speed, and executive functions (far transfer), but also on specific WM mechanisms discussed in the literature to potentially underlie such transfer effects. We found substantial to very strong evidence for the lack of near and far transfer effects as well as mechanism-specific effects, suggesting that neither WM capacity nor efficiency was enhanced through five weeks of intensive WM training. Instead, our findings led us to the conclusion that participants acquired task- and material-specific strategies. An alternative explanation for the absence of transfer, are ego-depletion effects that might have been emerged during the completion of the large test battery at pre- and posttest.

In the second study (No Consistent Evidence for Ego-Depletion Effects across Multiple Cognitive Tasks and Domains) we tested two hypotheses derived from the strength-model of self-control: (1) performance is worse at the end in comparison to the beginning of a series of tasks, (2) ego-depletion effects concern all tasks that require self-control. For this purpose, we reanalyzed data sets stemming from large cognitive test batteries administered in four previously reported studies. In these studies, participants were randomly assigned to one of two task orders, hence half of the participants completed the test batteries in reverse order than the remaining ones. Based on Bayesian inference, we had to reject both hypotheses, with the majority of Bayes factors providing at least substantial evidence in favor of null effects. These findings render it unlikely that ego-depletion is responsible for the lack of transfer in the first study reported in this thesis.

4.2 Future Directions

The main goal of this thesis was to investigate the specific mechanisms and processes that drive training and transfer gains in working memory. Additionally, we were interested in whether ego-depletion effects might counteract training and transfer gains as suggested by Green and colleagues (2014).

Our first study showed compelling evidence for the absence of any transfer after intense cognitive training. Moreover, both updating and binding group showed comparable performance in the specific WM mechanisms (i.e., focus switching, removal, and interference resolution) after training. Nonetheless, participants showed decisive evidence for a considerable performance increase in the trained tasks. As neither of the mentioned WM mechanisms is improved through training the question remains: what is the reason for the found training effects? The result pattern suggests that participants developed task-specific strategies during the training. This assumption is supported by the large number of participants, who retrospectively declared the usage of strategies during training. Reports of the applied strategies revealed that participants often used task- and material specific strategies, which could hardly be transferred to other tasks, thus it is not surprising that no transfer effects have been found. However, these results correspond to decades of research in skill acquisition, which revealed that expertise is highly specific and rarely transfers to other tasks (cf. Ericsson et al., 1980; Healy, Wohldmann, Sutton, & Bourne, 2006; Lewandowsky & Thomas, 2009).

As WM training not reliably generates the desired outcomes (i.e., transfer to other higher-order cognitive functions such as reasoning and executive functions), it might be reasonable to reconsider the previous approach in cognitive training research. Instead of just indirectly target a specific ability (e.g., reasoning) through WM training, one should specifically practice the ability that has to be improved (see also Melby-Lervåg et al., 2016). Since, the current study showed that participants strongly rely on strategies, it would be interesting to systematically investigate the kind of strategies participants applied in each task and how successful they were. Furthermore, one could examine whether the usage of specific strategies could predict training performance. In this context it would be worthwhile to examine whether there are task-independent strategies that target specific WM mechanisms and if so, whether they can be instructed to participants.

In our second study, we found at least substantial evidence for the absence of ego-depletion effects in extensive test batteries that lasted up to 4.5 hours. Thus, in contrast to the assumption of Green et al. (2014), it is rather unlikely that spending several hours on cognitive tests lead to ego-depletion effects. However, the absence of ego-depletion effects might have been caused by learning processes that counteracted ego-depletion. For example, Carter and McCullough, (2013a) found improved self-regulation after self-control exertion. This is in line, with research on learned industriousness (e.g. Eisenberger, 1992) and adaptation-level theory (Helson, 1964) that suggested adaption effects after self-control effort together with greater self-control exertion and, hence, increased performance. Therefore, future research should focus on the disentanglement of training and practice effects on the one side and ego-depletion effects on the other side.

4.3 Concluding Remarks

The empirical studies reported in this thesis contribute to the ongoing debate about the existence of training induced transfer effects and the malleability of higher-order cognitive functions such as WM. The first study revealed strong evidence for the absence of generalization after WM training and, thus, suggests that WM is a stable trait that cannot be fundamentally altered through training interventions. Given the large economic success of cognitive training interventions (SharpBrains, 2013, 2015), it is important to deliver a realistic view about their effectiveness. The second study showed that ego-depletion effects are not as common as expected in cognitive tasks. People are able to complete cognitively demanding tasks over a long period, without severe performance impairments.

5 References

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*, *22*, 366–377. http://doi.org/DOI 10.3758/s13423-014-0699-x

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*, 158–175. http://doi.org/10.3758/s13428-011-0123-7

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, 307–337.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100. http://doi.org/10.1037/0096-3445.133.1.83

Baumeister, R. F. (2014). Self-regulation, ego depletion, and inhibition. *Neuropsychologia*, *65*(2014), 313–319. http://doi.org/10.1016/j.neuropsychologia.2014.08.012

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*, 153–158. http://doi.org/10.1016/j.jesp.2016.02.003

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological Inquiry*, *7*(1), 1–15. http://doi.org/10.1207/s15327965pli0701_1

Baumeister, R. F., & Vohs, K. D. (2016). Misguided effort with elusive implications. *Perspectives on Psychological Science*, *11*(4), 574–575. http://doi.org/10.1017/CBO9781107415324.004

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351–355. http://doi.org/10.1111/j.1467-8721.2007.00534.x

Bellander, M., Eschen, A., Lövdén, M., Martin, M., Bäckman, L., & Brehmer, Y. (2017). No evidence for improved associative memory performance following process-based associative memory training in older adults. *Frontiers in Aging Neuroscience*, *8*, 1–11. http://doi.org/10.3389/fnagi.2016.00326

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: further evidence for cognitive plasticity in attentional control in late adulthood. *Experimental Aging Research*, *34*, 188–219. http://doi.org/10.1080/03610730802070068

Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., & Schwarz, N. (1994). Need for cognition: eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben: need for cogntion: a scale measuring engagement and happiness in cognitive tasks. *Zeitschrift Für Sozialpsychologie*, *25*, 147–154.

Bogg, T., & Lasecki, L. (2015). Reliable gains? Evidence for substantially underpowered designs in studies of working memory training transfer to fluid intelligence. *Frontiers in Psychology*, *5*, 1589. http://doi.org/10.3389/fpsyg.2014.01589

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, *8*, 445–454. http://doi.org/10.1177/1745691613491271

Borkenau, P., & Ostendorf, F. (2008). NEO-Fünf-Faktoren-Inventar nach Costa und McCrea (NEO-FFI). In *Manual* (2nd ed.). Göttingen: Hogrefe.

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, *6*, 1–7. http://doi.org/10.3389/fnhum.2012.00063

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*, 12–21. http://doi.org/10.1080/17470215808416249

Buschkuehl, M., M. Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience*, *2*, *Supplem*, S167–S179. http://doi.org/10.1016/j.dcn.2011.10.001

Button, K. S., Ioannidis, J. P. a, Mokrysz, C., Nosek, B. a, Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, *14*, 365–76. http://doi.org/10.1038/nrn3475

Carter, E. C., & McCullough, M. E. (2013). Is ego depletion too incredible? Evidence for the overestimation of the depletion effect. *Behavioral and Brain Sciences*, *36*, 683–684. http://doi.org/http://dx.doi.org/10.1017/s0140525x13000952

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, *5*(July), 1–11. http://doi.org/10.3389/fpsyg.2014.00823

Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General*, *144*(4), 796–815. http://doi.org/10.1037/xge0000083

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, *17*, 193–199. http://doi.org/10.3758/PBR.17.2.193

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory taks. *Nature*, *386*, 604–608.

Converse, P. D., & Deshon, R. P. (2009). A tale of two tasks: reversing the self-regulatory resource depletion effect. *Journal of Applied Psychology, 94*(5), 1318–1324. http://doi.org/10.1037/a0014604

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: Oxford University Press.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. http://doi.org/10.1177/0956797613504966

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320*, 1510–1512. http://doi.org/10.1126/science.1155466

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, *23*, 720–30. http://doi.org/10.1037/a0014296

de Jonge, P., & de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. *Personality and Individual Differences*, *21*, 1007–1020. http://doi.org/https://doi.org/10.1016/S0191-8869(96)00161-4

De Simoni, C., Luethi, M. S., Oberauer, K., & von Bastian, C. C. (2017). No consistent evidence for ego-depletion effects across multiple cognitive tasks and domains. *Manuscript Submitted for Publication*.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*, 353–386. http://doi.org/10.1080/09541440244000328

Deci, E. L., & Ryan, R. M. (2015). Intrinsic motivation inventory. Retrieved from http://selfdeterminationtheory.org/questionnaires/10-questionnaires/50

Dewitte, S., Bruyneel, S. D., & Geyskens, K. (2009). Self-regulating enhances self-regulation in subsequent consumer decisions involving similar response conflicts. *Journal of Consumer Research*, *36*(3), 394–405. http://doi.org/10.1086/598615

Dorbath, L., Hasselhorn, M., & Titz, C. (2011). Aging and executive functioning: A training study on focus-switching. *Frontiers in Psychology*, *2*, 1–12. http://doi.org/10.3389/fpsyg.2011.00257

Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2015). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-015-0865-9

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-s). *Journal of Personality Assessment*, *91*, 166–174. http://doi.org/10.1080/00223890802634290

Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science*, *16*, 915–925. http://doi.org/10.1111/desc.12068

Dweck, C. S. (2000). *Self-theories: their role in motivation, personality, and development*.

Philadelphia, PA: Taylor & Francis.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory : A specific updating process. *Journal of Memory and Language*, *74*, 77–90. http://doi.org/10.1016/j.jml.2013.09.003

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating : An experimental decomposition and individual differences, *36*, 170–189. http://doi.org/10.1037/a0017891

Eisenberger, R. (1992). Learned industriousness. *Psychological Review*, *99*(2), 248–267. http://doi.org/10.1037/0033-295X.99.2.248

Ekstrom, R. B. R., French, J. J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton NJ Educational Testing Service*. http://doi.org/10.1073/pnas.0506897102

Engle, R. W., & Turner, M. L. (1987). Is working memory capacity task specific. *Bulletin of the Psychonomic Society*, *25*, 331.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182. http://doi.org/10.1126/science.7375930

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task, 143–149.

Estrada, E., Ferrer, E., Abad, F. J., Román, F. J., & Colom, R. (2015). A general factor of intelligence fails to account for changes in tests' scores after cognitive practice: A longitudinal multi-group latent-variable study. *Intelligence*, *50*, 93–99. http://doi.org/10.1016/j.intell.2015.02.004

Friedman, N. P., Miyake, A., Schmeichel, B. J., & Tang, D. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172–179. http://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2009). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology*, *137*, 201–225. http://doi.org/10.1037/0096-3445.137.2.201.Individual

Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C. M., & Morrissey, R. A. (2012). The future promise of Cogmed working memory training. *Journal of Applied Research in Memory and Cognition*, *1*, 214–216. http://doi.org/10.1016/j.jarmac.2012.07.003

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research*, *78*, 756–772. http://doi.org/10.1007/s00426-013-0535-3

Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: evidence for the absence of transfer. *Manuscript in Preparation*.

Guye, S., De Simoni, C., & von Bastian, C. C. (2017). Do individual differences predict change in cognitive training performance? A latent growth curve modeling approach. *Manuscript Submitted for Publication*.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion dffect. *Perspectives on Psychological Science, 11*(4), 546–573. http://doi.org/10.1177/1745691616652873

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*(4), 495–525. http://doi.org/10.1037/a0019486

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*, 179–185. http://doi.org/10.1038/nmeth.3288

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*, 2409–2419. http://doi.org/10.1177/0956797613492984

Healy, A. F., Wohldmann, E. L., Sutton, E. M., & Bourne, L. E. (2006). Specificity effects in training and transfer of speeded responses. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 534–546. http://doi.org/10.1037/0278-7393.32.3.534

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York, NY: Harper & Row.

Herkert, J. (2012). *Trainings- und Transfereffekte des Arbeitsgedächtnisses* (Unpublished master's thesis). University of Zurich, Zurich, Switzerland.

Hossiep, R., Hasella, M., & Turck, D. (2001). *BOMAT-advanced-short version: Bochumer Matrizentest*. Göttingen: Hogrefe.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. http://doi.org/10.1080/10705519909540118

Hussey, E. K., Isaiah Harbison, J., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., & Novick, J. M. (2016). Memory and language improvements following cognitive control training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*. http://doi.org/10.1037/xlm0000283

Inzlicht, M., Gervais, W. M., & Berkman, E. T. (2016). Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015, 1–30.

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences*, *18*(3), 127–133. http://doi.org/10.1016/j.tics.2013.12.009

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6829–6833. http://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N -back task as a working memory measure. *Memory*, *18*, 394–412. http://doi.org/10.1080/09658211003702171

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition, 42*, 464–480. http://doi.org/10.3758/s13421-013-0364-z

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence, 38*, 625–635. http://doi.org/10.1016/j.intell.2010.09.001

Jäger, A. O., Süss, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS-Test)*. Göttingen: Hogrefe.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York: NY: Oxford University Press.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion-is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science*, *21*(11), 1686–1693. http://doi.org/10.1177/0956797610384745

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of “executive attention”. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 749–777. http://doi.org/10.1037/0278-7393.32.4.749

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*, 978–990. http://doi.org/10.1111/j.1467-7687.2009.00846.x

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*, 317–324. http://doi.org/10.1016/j.tics.2010.05.002

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*, 781–791. http://doi.org/10.1076/jcen.24.6.781.8395

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433. http://doi.org/10.1016/S0160-2896(05)80012-1

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized cognitive training in cognitively healthy older adults: A systematic review and meta-analysis of effect modifiers. *PLoS Medicine*, *11*. http://doi.org/10.1371/journal.pmed.1001756

Lewandowsky, S., & Thomas, J. (2009). Expertise: Acquisition, limitations, and control. *Reviews of Human Factors and Ergonomics*, *5*, 140–165. http://doi.org/10.1518/155723409X448044.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*, 571–585. http://doi.org/10.3758/BRM.42.2.571

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766. http://doi.org/10.1016/j.jesp.2013.03.013

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, *20*, 135–41. http://doi.org/10.3758/s13423-012-0335-6

Linares, R., Borella, E., Lechuga, M. T., Carretti, B., & Pelegrina, S. (2017). Training working memory updating in young adults. *Psychological Research*, *0*, 0. http://doi.org/10.1007/s00426-017-0843-0

Luethi, M. S., Friese, M., Binder, J., Boesiger, P., Luechinger, R., & Rasch, B. (2016). Motivational incentives lead to an over-increase in lateral prefrontal activity after self-control exertion. *Social Cognitive and Affective Neuroscience*. http://doi.org/10.1093/scan/nsw073

Lurquin, J. H., Michaelson, L. E., Barker, J. E., Gustavson, D. E., von Bastian, C. C., Carruth, N. P., & Miyake, A. (2016). No evidence of the ego-depletion effect across task characteristics and individual differences: A pre-registered study. *Plos One*, *11*(2), e0147770. http://doi.org/10.1371/journal.pone.0147770

Martinussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairment in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolsescent Psychiatry*, *44*, 377–384.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291. http://doi.org/10.1037/a0028228

Melby-Lervåg, M., Redick, T., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512–534. http://doi.org/10.1177/1745691616635612

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. http://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a H., Howerter, A., & Wager, T. D. (2000a). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, *41*, 49–100. http://doi.org/10.1006/cogp.1999.0734

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a H., Howerter, A., & Wager, T. D. (2000b). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, *41*, 49–100. http://doi.org/10.1006/cogp.1999.0734

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140. http://doi.org/10.1016/S1364-6613(03)00028-7

Morey, R. D. (2015, January 30). On verbal categories for the interpretation of Bayes factors

[Blog post]. Retrieved from http://bayesfactor.blogspot.ch/2015/01/on-verbal-categories-for-interpretation.html

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. Retrieved from https://cran.r-project.org/package=BayesFactor

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*, 46–60. http://doi.org/10.3758/s13423-010-0034-0

Murtagh, A., & Todd, S. (2004). Self-Regulation: A challenge to the strength model. *Journal of Articles in Support of the Null …*, *3*(1), 19–50. http://doi.org/10.1.1.387.3196

Navon, D. (1977). Forest before trees: The precedence of global fetures in visual perception. *Cognitive Psychology*, *9*, 353–383.

Noack, H., Lövdén, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. *Psychological Research*, *78*, 773–789. http://doi.org/10.1007/s00426-014-0564-6

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience*, *27*, 435–453. http://doi.org/10.3233/RNN-2009-0496

Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, *134*, 368–387. http://doi.org/10.1037/0096-3445.134.3.368

Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 197–214. http://doi.org/10.1037/0278-7393.32.2.197

Oberauer, K. (2008). How to say no: Single- and dual-process theories of short-term recognition tested on negative probes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*, 439–459. http://doi.org/10.1037/0278-7393.34.3.439

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 45–100).

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science*, *21*, 164–169. http://doi.org/10.1177/0963721412444727

Oberauer, K., & Lewandowsky, S. (2016). Control of information in working memory: Encoding and removal of distractors in the complex-span paradigm. *Cognition*, *156*, 106–128. http://doi.org/10.1016/j.cognition.2016.08.007

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language*, *51*, 80–96. http://doi.org/10.1016/j.jml.2004.03.003

Oberauer, K., Süss, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory. *Intelligence*, *31*, 167–193. http://doi.org/10.1016/S0160-

2896(02)00115-0

Oberauer, K., Süss, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence, 36*, 641–652. http://doi.org/10.1016/j.intell.2008.01.007

Oberauer, K., Süss, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in Working Memory* (pp. 49–75). Oxford: Oxford University Press.

Oelhafen, S., Nikolaidis, A., Padovani, T., Blaser, D., Koenig, T., & Perrig, W. J. (2013). Increased parietal activity after training of interference control. *Neuropsychologia*, *51*, 2781–2790. http://doi.org/10.1016/j.neuropsychologia.2013.08.012

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. Retrieved from URL: https://www.R-project.org/

Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Okford, England: Oxford psychologists Press.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*, 359–379. http://doi.org/10.1037/a0029082

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. *Diagnostica*, *47*, 57–66. http://doi.org/http://doi.org/10.1026//0012-1924.47.2.57

Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, 839–849. http://doi.org/10.1017/CBO9781107415324.004

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, *99*(3), 400–406. http://doi.org/10.1037/0033-2909.99.3.400

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. http://doi.org/10.3758/PBR.16.2.225

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences*, *508*, 297–306. Retrieved from http://psychsocgerontology.oxfordjournals.org/content/50B/6/P297.short

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General*, *136*(2), 241–255. http://doi.org/10.1037/0096-3445.136.2.241

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, *85*(1), 33–46. http://doi.org/10.1037/0022-3514.85.1.33

Schmiedek, F., Hildebrandt, A., Lövdén, M., Lindenberger, U., & Wilhelm, O. (2009). Complex span versus updating tasks of working memory: the gap is not that deep. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 1089–1096. http://doi.org/10.1037/a0015730

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*, 1–10. http://doi.org/10.3389/fnagi.2010.00027

Schmiedek, F., Oberauer, K., Wilhelm, O., Süss, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J Exp Psychol Gen*, *136*, 414–429. http://doi.org/10.1037/0096-3445.136.3.414

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*, 138–166. http://doi.org/10.1080/00461520.2015.1036274

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). Windsor, UK: NFER-NELSON.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Working memory training remains a work in progress. *Journal of Applied Research in Memory and Cognition*, *1*, 217–219. http://doi.org/10.1016/j.jarmac.2012.07.009

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, *17*, 245–276. http://doi.org/10.3758/PBR.17.2.193

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628–654. http://doi.org/10.1037/a0027473\r10.1037/0033-2909.131.1.30

Simon, R. J. (1990). The effects of an irrelevant directional cue on human information processing. *Advances in Psychology*, *65*, 31–86. http://doi.org/10.1016/S0166-4115(08)61218-2

Simon, J. R., Sly, P. E., & Vilapakkam, S. (1981). Toward the stimulus source. *Acta Psychologica*, *47*, 63–81.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training"; Programs Work? *Psychological Science in the Public Interest*, *17*, 103–186. http://doi.org/10.1177/1529100616661983

Sobel, K. V, Gerrie, M. P., Poole, B. J., & Kane, M. J. (2007). Individual differences in working memory capacity and visual search: the roles of top-down and bottom-up processing. *Psychonomic Bulletin & Review*, *14*, 840–845. http://doi.org/10.3758/BF03194109

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-016-1217-0

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., … Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence*, *41*, 638–663. http://doi.org/10.1016/j.intell.2013.07.013

Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science*, *25*(April), 1227–1234. http://doi.org/10.1177/0956797614526415

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology (2006)*, *59*, 745–759. http://doi.org/10.1080/17470210500162854

Stroop, R. J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643.

Süss, H. M., Oberauer, K., Wittman, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—And a little bit more. *Intelligence*, *30*, 261–288.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology. Human Perception and Performance*, *37*, 137–151. http://doi.org/10.1037/a0020365

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, *8*, 247–261.

Titz, C., & Karbach, J. (2014). Working memory and executive functions: effects of training on academic achievement. *Psychological Research*, *78*, 852–868. http://doi.org/10.1007/s00426-013-0537-1

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 166–179. http://doi.org/10.1037/0278-7393.18.1.166

Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: how to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 1322–37. http://doi.org/10.1037/0278-7393.30.6.1322

Vohs, K. D., Baumeister, R. F., & Schmeichel, B. J. (2013). Erratum to “Motivation, personal beliefs, and limited resources all contribute to self-control.” *Journal of Experimental Social Psychology*, *49*, 184–188. http://doi.org/10.1016/j.jesp.2012.08.007

von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological Research*, *80*(2), 181–194. http://doi.org/10.1007/s00426-015-0655-z

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, *69*, 36–58.

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research*, 1–18. http://doi.org/10.1007/s00426-013-0524-6

von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, *41*, 611–624. http://doi.org/10.3758/s13421-012-0280-7

von Bastian, C. C., Locher, A., & Ruflin, M. (2012). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*, 108–115. http://doi.org/10.3758/s13428-012-0224-y

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages : A test of four hypotheses. *Journal of Experimental Psychology: General*, *145*(2), 246–258. http://doi.org/10.1037/xge0000120

von Bastian, C. C., Guye, S., & De Simoni, C. (2017). How strong is the evidence for the effectiveness of working memory training? In M. F. Bunting, J. M. Novick, M. R. Dougherty, & R. W. Engle (Eds.), *Cognitive and Working Memory Training: Perspectives from Psychology, Neuroscience, and Human Development*. New York: NY: Oxford University Press. *Chapter Submitted for Publication.*

Waris, O., Soveri, A., & Laine, M. (2015). Transfer after working memory updating training. *PLoS ONE*, *10*, 1–19. http://doi.org/10.1371/journal.pone.0138734

Webb, T. L., & Sheeran, P. (2003). Can implementation intentions help to overcome ego-depletion? *Journal of Experimental Social Psychology*, *39*(3), 279–286. http://doi.org/10.1016/S0022-1031(02)00527-9

Weicker, J., Villringer, A., & Thöne-Otto, A. (2015). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology*, *30*, 1–23. http://doi.org/10.1037/neu0000227

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 1057–1064. http://doi.org/10.3758/s13423-012-0295-x

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 1–22. http://doi.org/10.3389/fpsyg.2013.00433

Xu, X., Demos, K. E., Leahey, T. M., Hart, C. N., Trautvetter, J., Coward, P., … Wing, R. R. (2014). Failure to replicate depletion of self-control. *PLoS ONE*, *9*(10), 1–5. http://doi.org/10.1371/journal.pone.0109950

Zimmermann, K., von Bastian, C. C., Röcke, C., Martin, M., & Eschen, A. (2016). Transfer after process-based object-location memory training in healthy older adults. *Psychology and Aging*, *31*, 798–814. http://doi.org/10.1037/pag0000123

Acknowledgement

First of all, I would like to thank Dr. Claudia von Bastian, who gave me the opportunity to write my thesis in such an interesting topic. I am very grateful for your support, guidance and advice. You always motivated me and to work with you was an important inspiration. I am very thankful having you by my side during my doctorate. I learned a lot from you. In addition, you gave me the opportunity for research stays at the University of Boulder and Bournemouth University. That was a wonderful and unforgettable experience.

My gratitude goes to Prof. Dr. Klaus Oberauer for the great ideas and your helpful inputs during the last years. Thanks for sharing your experience and knowledge. It was a great pleasure to work with you.

I was very happy to have Prof. Dr. Florian Schmiedek as my second supervisor. Whenever we met; he was a great inspiration and I enjoyed the discussion with him. Thank you for your helpful inputs.

Another big thank you goes to all the people from the cognitive psychology lab of the University of Zurich who shared their expertise and experience with me. I would like to especially thank all office colleagues Lea Bartsch, Mirjam Fuhrer, Marcel Niklaus and Mirko Thalmann (in alphabetical order). I would also like to thank all the students and research assistants who supported me in conducting the studies reported in this thesis. You were of great assistance.

Last but not least, I don't want forget to thank my family and my partner Silvan for their unrestricted support in the not always easy time for me. I would certainly not have been able to finish my Phd without you.

CURRICULUM VITAE

CARLA DE SIMONI

Department of Psychology, Cognitive Psychology, University of Zurich Binzmuehlestrasse 14, Box 22, 8050 Zurich, Switzerland

E-Mail: carla.de.simoni@psychologie.uzh.ch

Phone: +41 44 635 74 28

Web: http://www.psychologie.uzh.ch/de/fachrichtungen/allgpsy/Team/de-Simoni.html

EDUCATION

since 2013	PhD student at the Department of Psychology (Cognitive Psychology) University of Zurich, Switzerland Supervisor: Prof. Klaus Oberauer, Co-Supervisor: Florian Schmiedek
2013	**M. Sc.** at the University of Zurich, Switzerland Psychology (major), Business administration (minor) Master-Thesis: "Einfluss von Bilingualismus auf exekutive Funktionen und Arbeitsgedächtniskapazität bei jungen Erwachsenen" [The Impact of Bilingualism on Executive Functions and Working Memory Capacity in Young Adults]
2010	**B. Sc.** at the University of Zurich, Switzerland Psychology (major), Business administration (minor) Bachelor-Thesis: "Der Einfluss von Bilingualismus auf die exekutiven Funktionen über die Lebensspanne" [The Impact of Bilingualism on Executive Functions Across the Lifespan]

PROFESSIONAL EXPERIENCE

2016	Research stay at Bournmouth University, Department of Psychology, Dr. Claudia von Bastian

2015	Research stay University of Boulder Colorado, Department of Psychology and Neuroscience, Dr. Claudia von Bastian
2012	Internship ETH Zurich, Chair for Research on Learning and Instruction, supervised by Dr. Daniela Nussbaumer
2011	Research Assistant at University of Zurich, Department of Psychology, Cognitive Psychology, Dr. Miriam Gade
	Internship at University of Zurich, Departement of Psychology, Neuropsychology, Dr. Gianclaudio Casutt

TEACHING

Master's degree

2014	Seminar: "Fitness fürs Gehirn" [Fitness for the brain], University of Zurich (together with Sabrina Guye)
2013-2014	Exercise to lecture: "Applied Data Analysis with SPSS", Lucerne University of Applied Sciences and Arts (together with Simona Seidmann)

Bachelor's degree

2010	Seminar: "Interaktives Proseminar", University of Zurich

SUPERVISION

Master Theses

2015	Clara Overkott (Co-Supervision), topic: "The Effects of Reactivating Long-Term Memory Representations on Working Memory Performance"

Bachelor Theses

2015	Giorgia Vollmeier, topic: "Exekutive Funktionen und deren Rolle bei einer Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung" [Executive Functions and their Role in Attention Deficit Hyperactivity Disorder]
2016	Liridona Hotti, topic: "Training Exekutiver Funktionen" [Training of Executive Functions]

Research Interns

2015	Samuel Winiger, Doris Pfister, Monalisa Stiefel

Student Research Assistants

2014-2015	Lynn Häsler, Aline Gaillard, Jessica Oschwald, Samuel Winiger, Aylin Yantaz, Amanda Planzer, Tina Schwärzler, Clara Overkott

PUBLICATIONS

Submitted

von Bastian, C. C., Guye, S., & **De Simoni**, C. (2017). How strong is the evidence for the effectiveness of working memory training? In M. F. Bunting, J. M. Novick, M. R. Dougherty, & R. W. Engle (Eds.), *Cognitive and Working Memory Training: Perspectives from Psychology, Neuroscience, and Human Development*. New York: NY: Oxford University Press. *Chapter Submitted for Publication*.

De Simoni, C., Luethi, M. S., Oberauer, K., & von Bastian, C. C. (2017). No evidence for depletion effects in a wide range of cognitive tasks. *Manuscript Submitted for Publication*.

De Simoni, C., & von Bastian, C. C. (2017). No Evidence for Effects of Updating and Binding Training on Working Memory. *Manuscript Submitted for Publication*.

Guye, S., **De Simoni, C.,** & von Bastian, C. C.(2017). Do individual differences predict change in cognitive training performance? A latent growth curve modeling approach. *Manuscript Submitted for Publication*.

CONFERENCE CONTRIBUTIONS

Talks

2015 **De Simoni**, C., & von Bastian, C. C. (2015). *Training and transfer effects of working memory updating.* Invited talk at the 14th biannual congress of the Swiss Psychological Society (SSP/SGP), Geneva, Switzerland.

2014 **De Simoni**, C., & Gade, M. (2014). *Assessing a bilingual advantage in a multilingual context – a Swiss study*. Talk held at the 56. Tagung experimentell arbeitender Psychologen (TeaP), Giessen, Germany.

Posters

2016 **De Simoni**, C., & von Bastian, C. C. (2016). *Process-specific effects of working memory updating training.* Poster presented at the 57th annual meeting of the Psychonomic Society (Psychonomics), Boston, USA.

2015 **De Simoni**, C., von Bastian, C. C., & Oberauer, K. (2015). *Measuring the narrow focus of attention.* Poster presented at the 56th annual meeting of the Psychonomic Society (Psychonomics), Chicago, USA.

2014 **De Simoni**, C., von Bastian, C. C., & Oberauer, K. (2014). *The structure of working memory and its relation to reasoning.* Poster presented at the Lizenziandinnen-, Masterstudierenden- und Doktorandinnen-Kongress (LiMaDoKo), Zurich, Switzerland.

2013	**De Simoni**, C. & Gade, M. (2013). *Assessing a bilingual advantage in a multilingual context.* Poster presented at the Lizenziandinnen-, Masterstudierenden- und Doktorandinnen-Kongress (LiMaDoKo), Zurich, Switzerland.
	De Simoni, C. & Gade, M. (2013). *Assessing a bilingual advantage in a multilingual context.* Poster presented at the ESCOP international workshop on bilingualism and cognitive control, Krakow, Poland.

RESEARCH GRANTS AND AWARDS

2016	Suzanne and Hans Biäsch Foundation for Applied Psychology «Eigenschaften einer adäquaten Kontrollgruppe für kognitive Trainingsstudien», CHF 35’000.-
2015	Travel Grant from the PhD Program of the Department of Psychology (UZH) for the research stay in Boulder, Colorado, 2015
2013	Semester Award for Master Theses of the Faculty of Philosophy, University of Zurich
	Poster award of the LizentiandInnen- und Doktorierenden-Kongress 2013, Zurich, Switzerland